University of Oslo
Department of Informatics

Spectral moments and linear models used for photo-acoustic detection of crude oil in produced water

Fredrik Vogel

26th April 2001

Contents

List of Tables

List of Figures

0.1 Acknowledgments

At last finished writing the thesis, time to start with the preface and acknowledgments, as it ended up. I first want to recollect parts of the process leading forward to the final work. First of all the data set that I have been working on, was eight months late. The huge amount of information (480MB) was badly documented, the data had to be massaged and information combined from different files, before ending up with an acceptable data material. That was some of the bad experience.

Of the good experience is all the people that have been willing to help me forward with their knowledge and support and made it all possible. I want to thank Sverre Holm as my supervisor in Signal Processing, without his advise and knowledge the thesis would never have been where it is today. Ole Christian Lingjærde for his supervision and knowledge in statistics, always willing to talk and share his knowledge. To Kværner for supplying the project and sharing information and data, Tone Schanke as head of the project at Kværner for always being helpful and supplying the necessary information.

My Family for always being there. Katrine for her support through the final stages. Ann Philips for reading correction on my (Nor)English. USIT (Universitetets senter for Informasjons teknologi) for letting me use their office space and equipment as a part time employee. To RF (Realistforeningen): Skål kamerater... And everybody standing me close, for being my good friends.

0.2 Abstract

In this thesis, signal processing is performed on the output from the pulsed laser photoacoustic instrument monitoring crude oil in water. The instrument is constructed to perform inline monitoring of produced water in the pipeline during production. It is highly sensitive and testing was performed with hydrocarbons in water with concentrations in the range 0 - 1200 parts per million (ppm).

The thesis discusses the basic theory behind photoacoustic, and the construction of the instrument. Data material acquired during testing of the instrument is explored to improve the accuracy of the instrument.

The oil concentration is known to be affected by the following variables: photoacoustic response, temperature, salinity and pressure. These variables are analysed with statistical regression to show the instrument's ability to calibrate a specific compound crude oil.

Different methods of signal processing are used to enhance the result. When filtering, linear phase is necessary to avoid amplitude distortion of the peaks in the signal. This led to the use of a technique called spectral moments, a method that works directly on the Fourier spectrum and is insensible to phase. Statistics show that the spectral moments are able to enhance the result when equating the oil concentration. A new method for equating the oil concentration by filtering and arithmetic mean of the signal is discussed. The method is linked to the spectral moments with Parseval's theorem, it is easy to implement and statistics show good performance.

The thesis points out that the problems with a fouled instrument window must be solved to get the accuracy of the instrument down to the expected $\pm$ 100 ppm.

Chapter 1

Introduction

1.1 Background

The production of oil and gas is one of the largest industries in the world today. Crude oil is refined and used in combustion engines and is the key component in products made of plastic.

In the North Sea, oil platforms are a necessity in production of oil. Research has made it possible to retrieve oil at even greater sea depths. The construction of oil platforms is expensive, and can render a potential oilfield unprofitable.

During the production of oil and gas significant amounts of water are pumped up to the installation. The oil, gas and water solution is then put in a separation tank, and each part is extracted. The oil and gas is transported to shore, and the water is discharged to sea. This discharge is called produced water and contains approximately 1-2 % of hydrocarbons along with various levels of dissolved hydrocarbons, sediments, heavy metals, dissolved gases such as carbon dioxide and a number of chemicals used during the production process.

The monitoring of crude oil is important and necessary because environmental legislation demands such monitoring, but also because of the importance of keeping control over the production process. Methods already exists for oil in water monitoring. These methods require the aid of a person to withdraw a sample from the production line to be analysed further. The traditional method uses chemicals and solvents to analyse the sample, it takes time and is expensive both in labor and chemicals. Other methods involve looking at the absorption from different light spectra. All these methods have individual drawbacks and disadvantages, and are dependent on analysing a sample in the laboratory. Today there are no good methods for inline oil in water monitoring. The main object is to find a method that satisfies the demands of both environmental laws and the oil industry.

New technology makes it possible to move the oil production to platforms installed at the sea floor, thus cutting the construction expenses, making it possible to develop new oilfields. This type of platform is called a downhole installation. The name originates from the fact that the oil separation process is performed in the bore hole. In a downhole installation every process must be fully automated. These are some of the reasons the oil industry is on the lookout for a stable system that is able to do inline monitoring of produced water.

A promising and suitable instrument, using photoacoustic, was under development at Heriot-Watt University in Edinburgh. Kværner joined in the research and further development of a prototype at the request of Norsk Hydro. The prototype researched in this thesis is constructed to accommodate specifications needed to monitor produced water from a topside oil-platform, but a downhole version is also considered. At the request of Kværner the section of signal processing at the university of Oslo was asked to contribute in the project. After a two week period of testing in the first quarter of 2000, data was collected to analyse the performance of the prototype Oil In Water Monitor (OIWM). The main issue was to find an equation that can be used to compute the oil concentration from the parameters influencing the instrument response. The result of this analysis is found in chapter 5. The rest of the thesis is focusing on the signal produced from the instrument. New methods of analysing the signal are proposed in chapter 6. The Fourier spectrum illustrates the important components in the signal. Different types of filters are constructed and tested. Another method that is used with success is the spectral moments. The different methods are then analysed with statistics in chapter 7 to find the best one. The result is a simple to implement and efficient method that is a combination of filtering and the spectral moments.

1.2 Marine Pollution

The problem of pollution has become an issue affecting all humanity. As a result of environmental conferences and international agreements certain legislation exists around the legal amount of pollution that can be discharged into our environment. Such legislation must be followed by the oil producers in the North sea. The process of producing oil gives a certain amount of pollution, either accidentally or intentionally. An accident may occur during production that can result in high oil spills, and can have severe impact on the environment in both the short and long term. In the production of oil there is also some discharge into the sea that is accounted as normal. Such discharge can result in long term effects if not kept at a minimum and under control. Figure 1.1 shows the

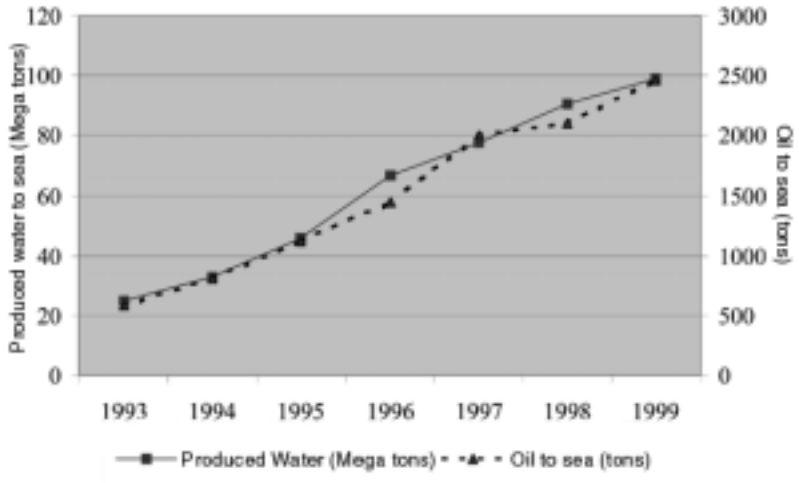

Figure 1.1: Correlation between discharge of oil and produced water [1].

discharge in the Norwegian sector in the years 1993 to 1999 [1]. The discharge of oil or oily water into the sea is prohibited by an international agreement called the Prevention of Oil Pollution Act 1971 [2]. There are certain exemptions given to this act for oil producing installations. The Oslo and Paris Commissions (OSPARCOM) has suggested the value for legal discharge set at 40 milligrams of oil per liter (mg/l) should not be exceeded over a monthly average. This control is done with lab measurement twice daily, no more than 4% of the samples each month must exceed $100 mg/l$. The term more widely used to describe oil content is part per million (ppm). Due to the relative density of oil and water the concentration will be approximately the same in both units. A topside installation can per legislation discharge produced water that has an oil content up to 100 ppm, compared to 200 ppm for downhole installations. The reason for this is that the discharge is happening further away from the sea surface, and has less impact on the environment.

Chapter 2

History and theory of photoacoustics

2.1 History of photoacoustics

The presented history of photoacoustic is an extract from a book written by Rosencwaig [4] in 1980. The photoacoustic effect goes back to Alexander Graham Bell who first observed it while working on a communication device in 1880. He was working on a way to transmit sound without any cables. By intensity modulating a light beam he used a selenium cell to pick up the change in light intensity and converted it to audible signals. He made the discovery that the signal could be attained directly without the electrical equipment. If the light was rapidly interrupted and focused on a solid (selenium), an audible signal could be picked up through a hearing tube. These discoveries were then further investigated and published in 1881. Similar experiments on gases were performed by John Tyndall and Wilhelm Röntgen in 1881 after hearing about Bell's discovery. Then the field lay dormant for 50 years until the discovery of the microphone made it possible to enhance the measurements. In 1938 Viengrov at the State Optical Institute of Leningrad used the method to study infrared absorption in gases and the gas content in gas mixtures. Pfund developed a gas analyser in 1939 in use at John Hopkins Hospital in Baltimore to measure CO and CO_2. Luft developed a commercial gas analyser which became available in 1946. The interest in photoacoustics grew in this period and it was only used to monitor gases. When the infrared spectrometer was invented, a more accurate method for monitoring gases existed so the field again lay dormant until 1970. It was the invention of the laser that gave the field of photoacoustics new possibilities. The effect was earlier called optoacoustic, but the name was changed to photoacoustic by Rosencwaig. This was to avoid confusion with the term acousto-optic effect which refers to another physical

phenomenon. The coherent, high optical energy in lasers was highly efficient for use in photoacoustics. The first use of photoacoustics with lasers was reported by White on solids in 1963, and Askar'yan et al. on liquids in 1964. New theories and understanding of photoacoustics was developed. This has led to quite an active field, today photoacoustics is being used in many different applications, and research in new areas is being done.

Photoacoustics is used in medical applications to measure glucose in samples [5], phantoms and human blood, it has been researched as a way to detect cancer, determination of melanin in human hair. Much research is performed in the detection of oil content in water. Most of this research has taken place at Heriot-Watt University in Edinburgh Scotland. Some Articles and Three Ph.Ds have been written about the subject; Hodgson in 1994 [6], Freeborn in 1997 [3], and Hannigan [7] in 1999.

2.2 The physics behind photoacoustic

Much theory has been developed to explain the phenomenon of photoacoustics. This section presents the fundamental elements in understanding the physic behind photoacoustics. High energy light is used, usually a laser, firing into the test sample. When the laser hits the sample, some of the energy is absorbed by the molecules in the media resulting in a region of higher temperature. The rise in temperature will generate an expanding region and a pressure wave will propagate away from the source. This decaying pressure wave is then picked up by a piezoelectric ceramic transducer. The phenomenon is contributed to the fact that the molecules in the sample being monitored have a higher response to light than the surrounding media. The process is pictured in figure 2.1.

The rest of the thesis will focus on photoacoustic propagation of hydrocarbons in water with special focus on the generated signal. There is one important formula that governs photoacoustics, and various authors have presented versions of this formula in their work. They have introduced particular assumptions and approximations which are relevant for their specifications. The formula used here is presented in the work of Freeborn [3, 8]. His formula looks especially at photoacoustics used in water, and the special cases that arise in this application.

$$P(r) = kE_0 \frac{\alpha\beta\nu^{\frac{1}{2}}}{c_p t_e^{\frac{3}{2}} r^{\frac{1}{2}}} \tag{2.1}$$

In formula 2.1 the magnitude of the pressure pulse is expressed by $P(r)$

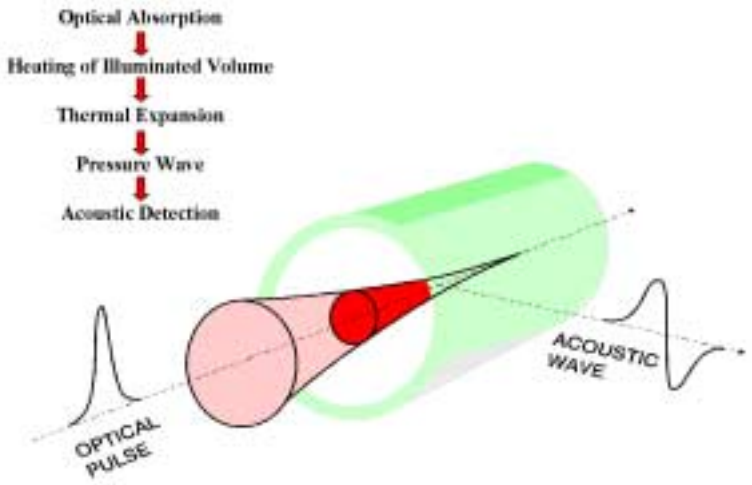

Figure 2.1: Principle of Photoacoustic propagation [18].

at a distance r from the optic axis, where E_0 is the incident optical energy, β is the volumetric thermal expansion coefficient, ν is the acoustic velocity, c_p is specific heat at constant pressure, k is a system constant and t_e is the effective time parameter. It relates the optical pulse width t_p and the acoustic transit time across the optical beam radius R, and $t_a = R/\nu$, such that $t_e = (t_p^2 + t_a^2)^{1/2}$. The formula 2.1 assumes relative weak absorption, as applies to water. That means that the optical absorption coefficient α and the optical path length l will define the magnitude of the acoustic region, and assume $\alpha l \ll 1$. In the weak absorption case the geometry of the acoustic signal can be regarded as cylindrical. It is also assumed that there is no thermal diffusion in the sample. The photoacoustic response, that makes it possible to monitor oil level in water relates to the high values of β and smaller values of c_p in crude oil compared to water.

Chapter 3

Instrumentation and test site environment

3.1 Instrumentation

The Oil In Water Monitor (OIWM) that is analysed in this thesis is a prototype developed by Kværner Oilfield Products together with Heriot-Watt University at the request of Norsk Hydro. The goal of the tests was to collect enough information and show that it is possible to construct a OIWM that can be put into production. Kværner has done the data collection and field-testing [12] of the prototype instrument at the Norsk Hydro research center in Porsgrunn. The OIWM prototype can be split into three main parts, the Head Unit, the Control Unit , and the User Interface PC.

3.1.1 Head Unit

The head unit is a housing that is capable of withstanding 100 bar of pipeline pressure. Light from the laser diode drivers in the control unit is transferred by a fiber-optic cable. The laser then enters the sample through a sapphire window. Below the sapphire window a half moon shape is protruding, this is the location of the Piezoelectric ceramic Transducer (PZT) used for acoustic detection. The positioning is illustrated in figure 3.1. The PZT detects photoacoustic pressure pulses in the range of $0-0,5Pa$ at ambient temperatures. The PZT output voltage is in the area $5-10uV$. The PZT is glued to a lead backing that dampen vibrations, but still keeps a strong photoacoustic response. Within 5 cm of the PZT a pre-amplifier card is placed to reduce attenuation. This amplifies the signal $62dB$, and result in a output signal of $40mV$ to $400mV$. A pressure transducer is mounted on a flange, and is expected to monitor pressure levels less than 100 bar, in a process environment

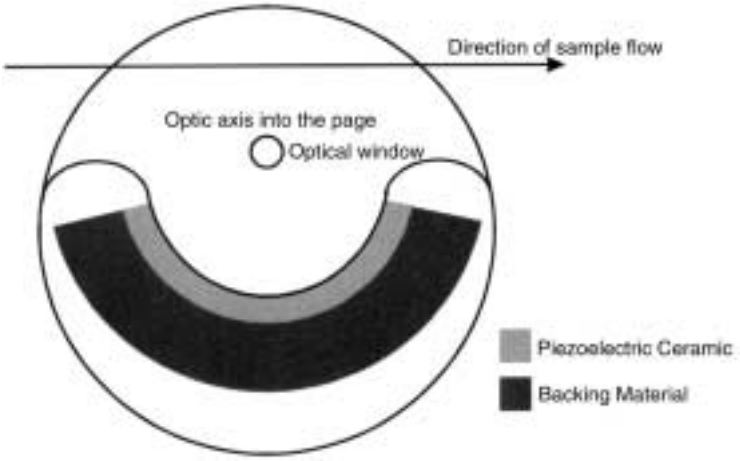

Figure 3.1: Front view of the head unit [3].

up to $100°C$. Part of the light is directed to an energy monitor with a beamsplitter, this way the laser energy and eventually any damage on the fiber-optic cable can be detected. The head unit includes a temperature sensor fastened to the housing with temperature conductive epoxy, for monitoring of pipeline temperature.

3.1.2 Control Unit

The Control Unit is connected to the head unit by an armored cable, through this cable laser pulses and information from the sensors in the head unit passes. The control unit contains two pulsed diode lasers, one with wavelength 905nm and another with wavelength 1550nm, firing sequentially. The light is transferred to the head unit through a fiber-optic cable. Due to losses in the system (couplers and fiber) only 30% of the energy is emitted into the sample fluid. The control unit contains the power supply for laser and electronics, together with PC-hardware running DOS. It also contains the fast analog-to-digital (A/D) card with 8 bit resolution that samples the signal and stores it on local disk. A keyboard and monitor allows direct control of the data acquisition and logging. The Control Unit operates independently, some of the data is sent via a serial link to the user interface PC.

3.1.3 User Interface PC

The User Interface PC is a standard PC running windows 95. Specialized software is run and displays output from sensors. It indicates system status and display alarms when values reach a preset threshold.

3.2 Oil in water monitor test site environment

The OIWM was tested at Hydro research center in Porsgrunn during the first quarter of 2000. One of the goals was to find an equation that expresses the oil concentration with respect to the variables salinity, pressure and temperature, and to get an idea of the error in such a formula. The values tested against reflect the values expected to be found in produced water. The test procedure was as follows; In the first test period the aim was to get data of the oil concentration versus temperature and photoacoustic response. Three temperatures were used as basis , $30°C$, $40°C$ and $50°C$. At the same time the oil concentration was varied along the values 0 ppm, 300 ppm, 600 ppm, 900 ppm, and 1200 ppm. In the last test period salinity was tested with various temperatures and oil concentrations. Pressure was also tested in the range of $1 - 30bar$. No change in photoacoustic response was observed [18] as a function of pressure in this range and has therefore been omitted in further analysis of the dataset. The test variables and their values are displayed in table 3.1.

Test variable	Range	Unit
Oil concentration	0, 300, 600, 900, 1200	ppm
Temperature	30, 35, 40, 45	$°C$
Pressure	1, 10, 20, 30	bar
Salinity	30 ,38, 45, 48	g/l

Table 3.1: Test parameters.

The instrument head was inserted into an one inch pipeline, normal sea water with salinity 30 g/l was pumped through the system introducing oil and additional salt when necessary. A pressure valve was used to regulate the pressure. An oil dispenser was used to set the oil concentration. Figure 3.2 shows the OIWM installed in the water rig at Hydro's research center at Porsgrunn.

During testing one or two samples were retrieved from the pipeline during a datapoint, and the oil concentration was measured with infrared spectrophotometric method [10] (IR) at a later time. When two samples existed the mean of the two values was used. The IR-method follows the Norwegian standard that is currently in use at oil installations. The oil concentration is not known exactly, because both the oil dispenser and the IR-method include measure errors, the correlation between the two methods is shown in figure 3.3. At this point there is uncertainty as to which of the two methods gives the best value. Because of this uncertainty both the IR-method and the dispenser values are used when trying to find models expressing the oil concentration,

Figure 3.2: OIWM installed in the water rig at the Hydro's Research Center in Porsgrunn.

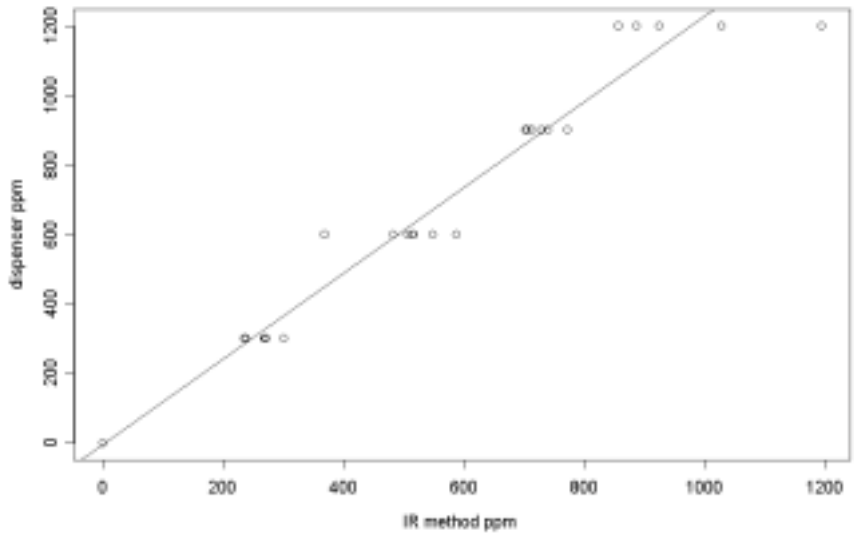

Figure 3.3: Plot of the IR-method versus the dispenser oil concentration.

each is presented with separate analysis. Any error in the reference method will be inherited by the models constructed from these measurements. Not all the data points were monitored with the IR-method, so less data is available in this case. The first week of testing Brage oil was used, this is a relatively thick oil and stuck easily to the OIWM window, resulting in measure errors in the form of heightened response. No results were obtained with this oil. Instead the thinner Visund oil was used in the rest of the period, giving better results, but still contributing to some problems. During a one day run the OIWM acted stably, but after running for several days the sapphire window became fouled and resulted in an increased response. The window was cleaned to get the response back to base level. This procedure was performed three times during the Porsgrunn testing.

3.3 Signal description

The control unit receives the amplified PZT signal from the instrument head. The signal is processed with a fast Analog to Digital card (A/D), with a sampling frequency at 100 MHz, and with a precision of 8 bit. The card acquires 1000 samples at intervals of 10 nano seconds. In earlier experiments amplitude analysis has been used to find the photoacoustic response. This value is found by looking at the difference between the highest and lowest peak of the signal within a specified window. This value is called the peak-to-peak value, a typical signal with the peak-to-peak value marked off is shown in figure 3.4.

3.3.1 Signal averaging

The signal generated from one pulse of the laser contains much noise, and it is nearly impossible to distinguish any special features. The noise is white, and comes from induction in cables, electronics and physical vibrations picked up by the transducer. To enhance the signal-noise ratio an averaging of signals is performed. Earlier testing [13] show that an average over 1500 pulses is reasonable for the 905nm wavelength, and 500 pulses for the 1550nm wavelength. It was believed that the accuracy could be further improved so the prototype uses averaging over 1500 pulses times 10. The instrument averages in steps of 1500 pulses, each time finding the peak-to-peak value of this average. The ten last peak-to-peak values are averaged and the value is written to the log file. This value is called a datapoint and consists of a total average of 15000 pulses. The waveform signal, as shown in figure 3.4, consists of the average of the last 1500 pulses of the ten and is also written to log file at this time. This means that the peak-to-peak value found from the waveform

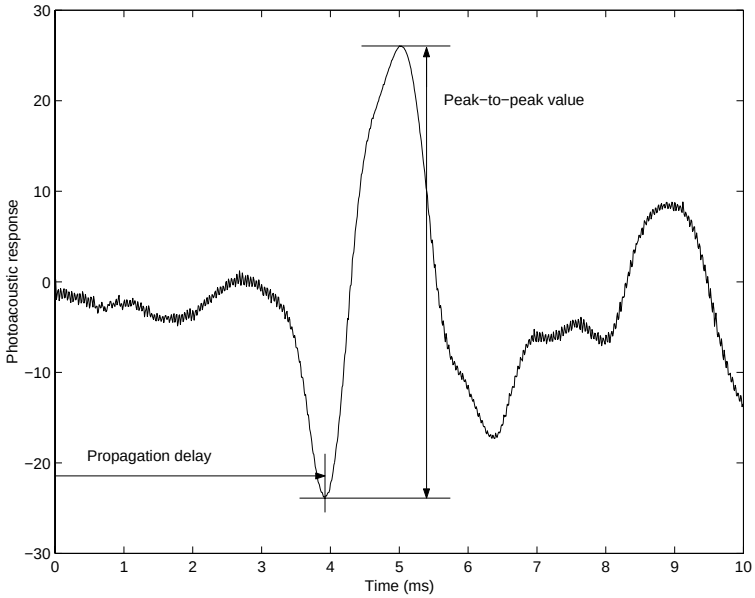

Figure 3.4: Typical photoacoustic signal from the OIWM.

signal consists of an average of 1500 pulses compared to a datapoint consisting of 15000 pulses. Because of this averaging process, the peak-to-peak value from a datapoint will have a better signal-noise ratio value compared to the peak-to-peak found from the waveform signal.

3.3.2 Physical dependencies

Earlier laboratory experiments [3] have shown that the photoacoustic response is connected to the level of salinity and temperature. This relationship follows from equation 2.1. The key physical parameters are the thermal expansion coefficient, β, specific heat capacity, c_p, and acoustic velocity, ν, all vary with both temperature and salinity. It is also known that equations expressing acoustic velocity in water are affected by pressure. Pressure was tested again at the Hydro research center at Porsgrunn to see if it affected the photoacoustic response, but no dependencies were visible in the range up to 30 bar [18].

Acoustic propagation delay

The time between delivery of the optical pulse and the detection of an acoustic signal is called the acoustic propagation delay. This delay, t_{pk}, is determined by the acoustic speed in water, ν and distance to the transducer, r. In addition comes a system constant that relates to the delay

in system electronics, t_{sys}. The relation is expressed in 3.1, and is displayed in figure 3.4.

$$t_{pk} = \frac{r}{v} + t_{sys} \tag{3.1}$$

The acoustic velocity in water is well documented and is known to only be related to the three physical quantities, temperature, salinity and pressure [22]. No other physical properties have been found to affect the velocity of sound in sea water. It is reasonable to believe that it is possible to find the salinity from the above mentioned parameters, and use this salinity when equating the oil concentration. This issue is researched further in section 5.6.

The effect of optical energy on signal response

When calculating the photoacoustic response, the optical energy entering the sample has significant effect. In laboratory tests the peak-to-peak signal was found to be directly proportional with energy [3]. The energy monitor in the head unit picks up the energy in incident optical light, this value is used to normalize the signal. In this way variations in optical energy can be largely disregarded. All values used in analyses in this thesis use the normalized signal.

The effect of crude oil on signal response

Crude oil is made up from a long range of different hydrocarbons, each type comes in different quantities, depending on the oil reservoir. The number of carbon atoms in a hydrocarbon molecule describe the molecular size. High quantities of large molecules in a sample, result in high photoacoustic response. The photoacoustic response of crude oil was tested against different wavelengths of light [3]. The difference in response between crude oil and salt water shows better characteristics at the 905 nm wavelength than the 1550 nm wavelength. The 1550 nm wavelength was choosed for the low response to crude oil to possible detect other effects. It was primary added to be researched as a base level detector that could be used to remove the effect of the oil film on the instrument window. This was not seen as an important issue by Kværner and is the reason the 905nm wavelength is the only light source used in further analyses.

Other test-parameters

If the instrument is to work in production all consideration must be accounted for. The performance when the produced water contain particles, chemicals, and gas, and the reaction towards flow rate and droplet size

encountered in the pipeline. These were important questions to answer before starting development of the prototype.

These parameters were tested at Norsk Hydro Research Center in Porsgrunn in the last quarter of 1998 [11]. The droplet size was tested in the range $20-30\mu m$ and no changes was recorded in the photoacoustic response. Theoretically the flow rate is not expected to change the photoacoustic response. The measurement can be seen as a snapshot within a period of $5\mu s$, in this period the sample will have moved $10\mu m$ which is small compared to the acoustic wavelength of around $1.5mm$ and the travel length to the PZT. The instrument insensibility to flow was proved in these tests.

To test the influence of gas, nitrogen was injected into the pipeline. The result shows a decrease in signal with increasing amount of gas. This can be the cause of scattering of the laser, or because nitrogen absorbs the wavelength. When the amount of nitrogen was kept constant, the photoacoustic response showed a constant linear trend, no further tests have been performed on this issue.

More tests were run with sand and Bentonite. Sand does not seem to change the result, while Bentonite showed a decrease in the photoacoustic response, few test points where run so it is hard to do draw a firm conclusion. Chemicals were run with methanol up to 3 %. Dissolved hydrocarbons were seen as one of the big challenges, but little influence on the photoacoustic response was recorded. The conclusion after these tests was that the important factors that influence the photoacoustic response are pressure, salinity and temperature.

Chapter 4

Methods of analysis of the photoacoustic signal

4.1 Spectral moments

The spectral moments is a method that retrieves information directly from the Fourier spectrum. It is advantageous when we wish to extract special features in the signal and is not dependent on keeping the original signal itself. One of the advantages is the insensitivity to changes in the phase of the signal. Spectral moments use the power of the Fourier spectrum when making calculations. Filters that distort the phase might ruin the signal, but the spectral components will stay untouched. The spectral moments are a method usually used in image processing [16], the central moments are derived from the spectral moments and work as a transformation that can mirror image, rotate and resize an image. The spectral moments can also be viewed as a way to use statistics to analyse the power spectrum. The different moments have a statistical interpretation that will be described below.

The one dimensional spectral moment [17], m_n, of size n of a continuous power spectrum $S(f)$ is defined in 4.1.

$$m_n = \int_0^\infty f^n S(f) df \tag{4.1}$$

When working on a discrete-time signal the spectral moment is defined in equation 4.2. The length of the power spectrum is N, and the sampling interval is Δt.

$$m_n = \frac{2}{N} \sum_{l=0}^{(N/2)-1} S(l)(l/N\Delta t)^n \tag{4.2}$$

We wish to limit our computations to a band of frequencies, this will correspond to a filtering process where the frequencies of interest are

extracted. In the continuous case we set the frequency limits of the integral from c_1 to c_2, and get equation 4.3

$$m_n = \int_{c_1}^{c_2} f^n S(f) df \tag{4.3}$$

In the discrete case the spectral estimate is done in the frequency range c_1 to c_2 of length I, where $I = c_2 - c_1$. The sampling interval is still Δt. The result is equation 4.4

$$m_n = \frac{1}{I} \sum_{l=c_1}^{c_2} S(l)(l/N\Delta t)^n \tag{4.4}$$

The zero order moment will in the discrete case look like equation 4.5.

$$m_0 = \frac{1}{I} \sum_{l=c_1}^{c_2} S(l) \tag{4.5}$$

The zero-order spectral moment is basicly the average of the values in the interval c_1 to c_2, and is proportional to the mean energy in that interval. If we use equation 4.2 to take the zero order spectral moment over the whole signal, it can be estimated directly from the variance from the time series itself. When only working on parts of the spectrum it is necessary to normalize the spectrum with m_0 before taking the higher spectral moments. This will level out the effect of the size of the values of higher order moments, and make it possible to extract information about the shape of the spectrum. The first order spectral moments in the continuous case used on part of the spectrum is defined in equation 4.6

$$m_1 = \frac{1}{m_0} \int_{c_1}^{c_2} f^n S(f) df \tag{4.6}$$

The discrete first order moment will then look like equation 4.7

$$m_1 = \frac{1}{I m_0} \sum_{l=c_1}^{c_2} S(l)(l/N\Delta t)^n \tag{4.7}$$

The first order spectral moment is interpreted as the mean frequency of the signal in the interval I. To get information about the higher order moments one needs to define the central moments 4.8.

$$M_n = \frac{1}{m_0} \int_{c_1}^{c_2} (f - \bar{f})^n S(f) df \tag{4.8}$$

The mean frequency, $\bar{f}$ is found from the 1. order moment. The discrete version of this formula is given by equation 4.9

$$M_n = \frac{1}{I m_0} \sum_{l=c_1}^{c_2} S(l)(\frac{l}{N\Delta t} - \bar{l})^n \tag{4.9}$$

Where the mean frequency $\bar{l}$ is found from the discrete 1. order moment. The second order central moment can be interpreted as the variance of the power spectrum, or the squared bandwidth. The third order is the skewness, and the fourth order is the kurtosis. The central moments can be easily found from the moments.

4.2 Linear regression analysis

A method that tries to fit an equation to a set of data is called regression. Among the various methods of regression, linear regression is the simplest, and the most widely used. Descriptions of the method are found in almost any elementary textbook in statistics [15]. In two dimensions a line may be described in algebraic terms on the form 4.10.

$$y = \alpha_0 + \alpha_1 x \tag{4.10}$$

In the real world, observations usually do not fit an equation like this, even if the true relationship is linear. A term expressing the noise in each observation must be added. The model then becomes

$$y_i = \alpha_0 + \alpha_1 x_i + \epsilon_i \tag{4.11}$$

where the subscript i in the equation stands for observation number i. The term ϵ_i is the noise in the observation. In general we may have more than one independent variable, and a multiple regression model such as

$$y_i = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \cdots + \alpha_k x_{ik} + \epsilon_i \tag{4.12}$$

may be considered. This model has $i = 1, \ldots, n$ observations and $k \leq 1$ independent variables. The predicted values $\hat{y}_i$ are expressed with

$$\hat{y}_i = \hat{\alpha}_0 + \hat{\alpha}_1 x_{i1} + \hat{\alpha}_2 x_{i2} + \cdots + \hat{\alpha}_k x_{ik} \tag{4.13}$$

where $\hat{\alpha}_0, \ldots, \hat{\alpha}_k$ is the estimate of $\alpha_0, \ldots, \alpha_k$ in equation 4.12. The residual error can then be defined as

$$e_i = y_i - \hat{y}_i \tag{4.14}$$

where y_i is the observation, and $\hat{y}_i$ is the predicted value found using formula 4.13. Equation 4.12 may be written in matrix notation as

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_k \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} \tag{4.15}$$

or simply as

$$\mathbf{y} = \mathbf{X}\alpha + \epsilon \tag{4.16}$$

where $\epsilon, \mathbf{y} \in \mathbb{R}^{n}$, $\mathbf{X} \in \mathbb{R}^{n,k+1}$, and $\alpha \in \mathbb{R}^{k+1}$. The goal is to find a method to estimate the parameters α_j. One way to obtain an estimate for α is by minimizing the square sum of the error.

$$\begin{aligned} S = \sum_{i=1}^{n} \epsilon_i^2 &= \sum_{i=1}^{n} (y_i - \alpha_0 - \alpha_1 x_{i1} - \alpha_2 x_{i2} - \cdots - \alpha_k x_{ik})^2 \\ &= (\mathbf{y} - \mathbf{X}\alpha)'(\mathbf{y} - \mathbf{X}\alpha) = \mathbf{y}'\mathbf{y} - 2\alpha'(\mathbf{X}'\mathbf{y}) + \alpha'(\mathbf{X}'\mathbf{X})\alpha \end{aligned} \tag{4.17}$$

In order to find the minimizer, we use matrix differentiation on equation 4.17 and get equation 4.18

$$\frac{\delta S}{\delta \alpha} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\alpha \tag{4.18}$$

Setting equation 4.18 equal to zero, we see that the least squares estimate satisfies

$$(\mathbf{X}'\mathbf{X})\alpha = \mathbf{X}'\mathbf{y} \tag{4.19}$$

If $\mathbf{X}'\mathbf{X}$ is a non-singular matrix, equation 4.19 has the unique solution

$$\alpha = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{4.20}$$

The vector α then gives the minimum of S in equation 4.17.

4.3 Variable selection

Frequently we start with a long list of independent variables that we suspect have some effect on the dependent variable, but for various reasons we want to reduce the list. To test the effect of the jth term the following hypothesis may be considered:

$$H_0 : a_j = 0 \qquad versus \qquad H_1 : a_j \neq 0$$

The decision we want to make is whether we should accept or reject the hypothesis that variable a_j is equal to zero. Using Student's t-test we consider the test statistics

$$t = \frac{\hat{a}_j}{se(\hat{a}_j)} \tag{4.21}$$

Which, under the assumption of normal errors, follows a Student's t-distribution with one degree of freedom under the hypothesis H_0 where se is the standard error of the independent variable. The $\hat{a}_j$ is the estimated value of the variable a_j. The p-value is defined below

$$p = prob(T > |\hat{t}_{obs}|) \tag{4.22}$$

where T follows a Student's t-distribution with one degree of freedom. The smaller the p-value, the more significant is the corresponding term in the model. The term $\hat{t}_{obs}$ indicates if a observed value in the model lie within or outside the estimated value in interval $(-\hat{t}, \hat{t})$ and $\hat{t} \sim t$. Under normal assumptions we will conclude that a parameter is significantly different from zero if the p-value is less than, lets say 0.05. Another method used to decide on variables and comparing models is Mallow's C_p, and is defined as:

$$C_p = \frac{RSS}{s^2} - (n - 2p) \tag{4.23}$$

Here RSS is the Residual Sum of Squares and is defined as $\sum_{i=1}^{n} e^2$ where e is the residual from equation 4.14. The term p is the number of independent variables, n is number of observations, and s^2 is an unbiased estimator of σ^2. The term s^2 is defined as

$$s^2 = \sum_{i=1}^{n} \frac{e_i^2}{n - k - 1} \tag{4.24}$$

where the term (n-k-1) is the degrees of freedom, where k is number of independent variables, and n is the number of observation.

The purpose of variable selection procedures is to select or help select from the total number of candidate variables to a smaller subset. A number of procedures can be performed, selecting the best candidate with either s^2,C_p or possible other types of statistic. In many situations there is rarely one obvious best equation, and the near winners are almost as good. One approach when searching for the best model is to inspect all possible subsets. Another method that consume less resources is a stepwise selection, most statistical programs have procedures for removing the less important variables one by one. The method of stepwise selection often removes important variables from a model. It will often be necessary to do manual work and it is advised that the analyst use his or her intuition.

4.4 Speed of sound in water

The equations used to measure the speed of sound in water include the three variables temperature, salinity, and pressure. No other physical quantities affecting the value are found. The cross correlation between the three terms appears to be complicated. Tables have been widely in use to find the relation, but around 1970 equations are able to express the value more accurately. Del Grosso [23] gives an equation containing 19 terms each to 12 significant figures in the powers and cross products

of the three variables. Lowett [24] presents a less cumbersome equation, but still values of produced water are far outside the values used in his dataset. H. Medwin has a simplified equation with a larger error with limits in an interesting range [22] as seen in equation 4.25

$$\begin{array}{lll} c = & 1449.2 + 4.6T - 5.5 \times 10^{-2}T^2 & 0 \leq T \leq 35^\circ \\ & +2.9 \times 10^{-4}T^3 + (1.34 - 10^{-2}T)(S-35) & 0 \leq S \leq 45 \\ & +1.6 \times 10^{-2}D & 0 \leq D \leq 1000 \end{array} \tag{4.25}$$

T is the Temperature in Celsius, S is salinity in parts per thousand, D is the depth in meters, and c is the velocity of sound in water. The depth variable in equation 4.25 corresponds to the pressure variable used in the OIWM analysis.

Chapter 5

OIWM Estimation of Parameters for Linear Models

5.1 Introduction

The instrument has been through a series of testing at the Norsk Hydro research center in Porsgrunn, the testing was performed in the first quarter of 2000. Kværner performed the final preparation of the data. The purpose of this chapter is to propose some linear models produced from the data that can be used to predict the oil concentration.

The photoacoustic response used in this analysis is the normalized peak to peak value. This is the value that is automatically equated by the OIWM during testing. To represent the oil concentration we use both the oil dispenser and the IR-measurements. We want to model the two on behalf of photoacoustic response, temperature and salinity. The test matrix and other circumstances around the testing are described in section 3.2.

5.2 The test setup

Prior to the Porsgrunn tests an older version of the instrument had shown good results in laboratory experiments both at Heriot-Watt university, and in field experiments. A strong linear relationship was proved between the photoacoustic response and the oil concentration. Linear dependencies towards temperature, salinity and pressure were also found. The interaction between each and all variables was not checked at this stage. The instrument used in earlier testing was replaced with a new prototype with enhanced electronics. At the request of Kværner a project to find an equation to express the oil concentration as a function of salinity, pressure, and temperature was assigned to the section

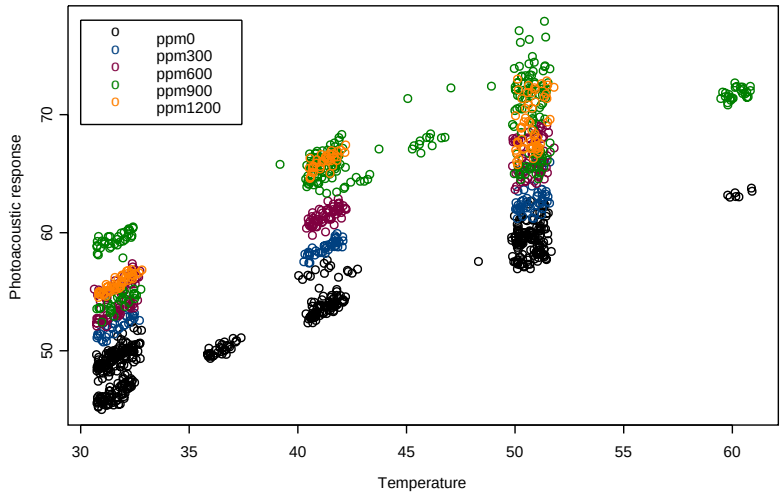

Figure 5.1: All data, using oil concentration from the oil dispenser.

of signal processing at the University of Oslo. Because of the linearity observed in prior experiments, it was assumed that a linear model would be sufficient to find the oil concentration. It was still a possibility of strong interaction between variables that would make a linear model inappropriate.

First let us have a look at the relationship between photoacoustic response and the temperature. Figure 5.1 is a plot of all the data acquired during testing, using the values from the oil dispenser for oil concentration. The temperature varies from $30°C$ to $60°C$. The oil concentration is varying in the area from 0 to 1200 ppm in steps of 300 ppm. The plot can be viewed as a three-dimensional figure, where the third axis is perpendicular to the paper. The oil concentration is colour coded to illustrate this effect. The plot exhibits the expected linear trend observed in earlier experiments. The effects of salinity and pressure are not accounted for in these plots. The results at higher oil concentrations might indicate that a nonlinear model will be more appropriate, however this is difficult to decide on the basis of the current data.

Figure 5.2 shows another way of presenting the data. The figure displays the oil concentration using the IR-method versus the photoacoustic signal for $30°C$, $40°C$ and $50°C$. The oil concentration from the IR-method varies, so the temperature is the fixed parameter instead of the oil concentration. The plot in figure 5.2 does not consider the temperature variations of $2-3°C$ at the different temperature levels. This

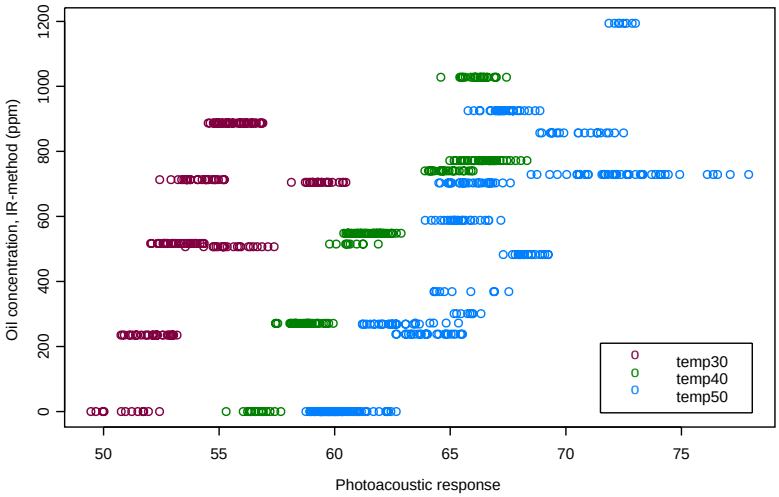

Figure 5.2: All data, using oil concentration from the IR-method.

temperature change gives at first glance at figure 5.1 a change of about 2-3 in photoacoustic response. This will influence the placement of the data. Samples of the IR-method were collected once or twice during a test point. When more than one sample were collected, the mean of the samples were used in further analysis to represent the oil concentration. The variations in the samples were often more than the ±10% that the IR-method is known to exhibit [25]. This can be the result of changes in values delivered by the oil dispenser, background noise or measurement errors. Little information exists about the general stability of the oil dispenser. The plot in figure 5.2 contains less data than figure 5.1, because the oil concentration was not checked with the IR-method at every run. In figure 5.2 the photoacoustic response relates to the IR-method at different temperatures. The IR-measurement is clustered in horizontal lines, where each line corresponds to one datapoint. This also illustrates the variations in the photoacoustic response within each datapoint. These variations can be due to change in temperature or differences in oil concentration delivered by the oil dispenser during a run. The plot shows that the IR-method relates linearly to the photoacoustic response within the three fixed temperatures displayed. From figure 5.1, and 5.2 it is believed that a linear model will be appropriate between the oil concentration and the temperature in further analyses. It is also believed that temperature is the variable that has the largest impact on the photoacoustic response.

5.3 Linearity in data

We start the analysis by exploring the linearity of a single level of oil concentration, first working on the data acquired from the oil dispenser. A separate analysis is performed on the five different oil concentrations. The level of salinity is kept constant at $30g/l$; this is the natural salinity of normal sea water. This means that the salinity predictor is removed and we are working on a smaller dataset than shown in figure 5.1. To ensure that the oil concentration used in the regression is independent variables we use the mean of the values within each datapoint. It is the true oil concentration we want to model, but we use the dispenser oil and IR-measurements to represent the value. It is this value that will be used in the analyses for the rest of this chapter. The reason is that the oil concentration, both the IR-method and the oil dispenser, is the same within a datapoint. This behavior can affect the output values of standard deviation and the p-value. Each datapoint is therefore counted as one observation. The total dataset is then reduced from 1607 observations to 84 observations. At this point it is still possible to visualise the data graphically. Linear regression is used to find a linear fit for each of the five oil concentrations. The five regression lines are displayed together with the data in figure 5.3. The figure shows that the photoacoustic response increases with increasing oil concentration. This is the same result that is observed in earlier experiments. At the higher oil concentrations inconsistency occurs, the oil concentration of 1200 ppm intersects the line of 900 ppm. This plot can help to understand and check the quality of the data. It is believed that many of the problems at higher oil concentrations arise from the fouling of the instrument window. Over time an oil layer builds up on the instrument window, resulting in a heightened photoacoustic response. Other sources of error can be bookkeeping or problems with the oil dispenser at higher oil concentrations. In the last case a better fit can be achieved with the IR-method. It is therefore interesting to look at the linearity of the data monitored with the IR-method. This analysis includes only the salinities of normal seawater, data with values outside this range is removed. Two datapoints were found to have salinity higher than normal seawater, and were removed. The oil dispenser method had eighteen such datapoints in comparison. Only two salinity values with concentration higher than seawater make a poor basis when trying to construct models that consider salinity as a predictor. The size of the dataset is more than halved because fewer datapoints was monitored with the IR-method. The method using the oil concentration from the dispenser has 66 observations in comparison to the IR-method with 31 observations. Many observations are usually advantageous when doing statistical analysis, helping towards a better and more accurate result. In this analysis

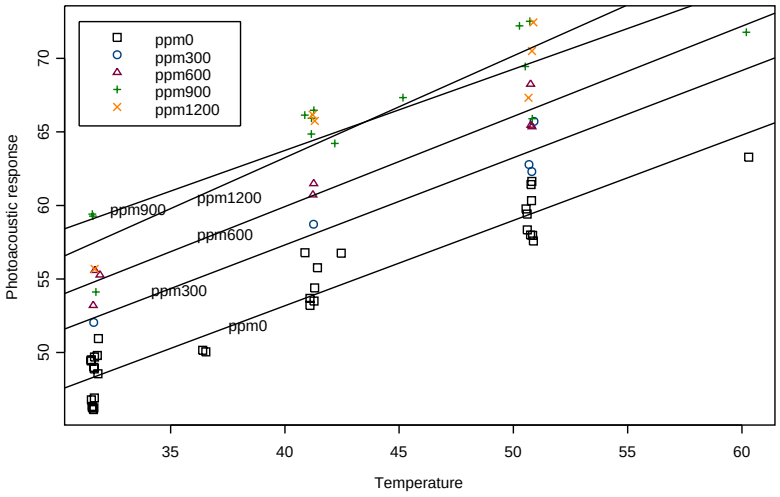

Figure 5.3: Data showing photoacoustic response versus temperature with corresponding regression lines.

we wish to display the relation between the photoacoustic response and oil concentration from the IR-method. Each of the three temperatures $30°C$, $40°C$, and $50°C$ is displayed in the same plot. We are using the mean of each datapoint with the same reasoning as mentioned previously. The plot is displayed in figure 5.4 and can be visualised as a three dimensional space with the colour-coded temperature perpendicular on the paper. Linear regression is used to fit the lines. The plot shows few available values appearing above the oil concentrations of 800 ppm, and the values appearing at this level display a bad fit to the regression line. This indicates that problems at higher oil concentrations might exist when using the IR-method.

5.4 Estimation in three dimensions

By constructing a model that includes the three most important predictors the oil concentration can be estimated as a function of photoacoustic response and temperature. Salinity is still excluded outside levels of normal seawater. First a model based on the dispenser oil concentration is investigated. The oil dispenser level is a set value with five levels, and not containing noise in the statistical sense. When constructing models based on the dispenser oil concentration we must be aware of this fact.

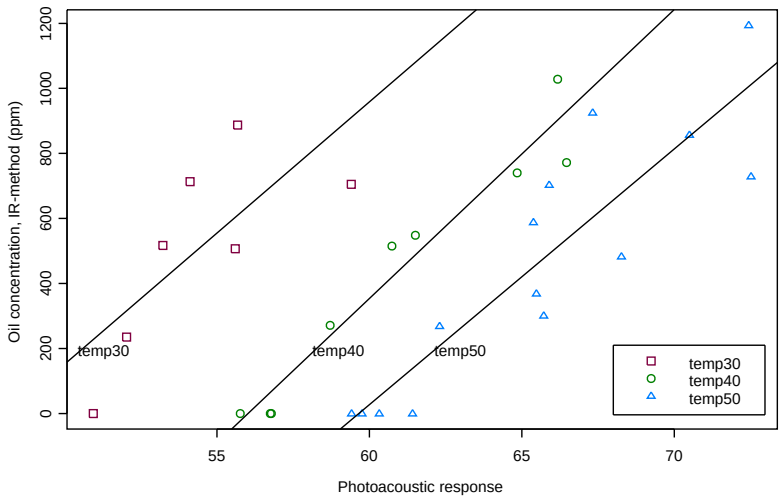

Figure 5.4: Data showing photoacoustic response versus IR-method oil concentration with corresponding regression lines.

The oil concentration is expressed with the two variables temperature and photoacoustic response, the model is displayed in equation 5.1.

$$oil_i = \alpha_0 + \alpha_1 pr_i + \alpha_2 temp_i + \epsilon_i \tag{5.1}$$

The oil_i is the true oil concentration, pr_i is the photoacoustic response of measurement number i, $temp_i$ is the temperature at the same instant. Since the true oil concentration is unobserved, the oil dispenser level is used as a proxy. Therefore the term ϵ_i incorperates observational error as well as modelling error. The error is due to differences in the set oil level and the true oil level. The model estimates the parameters α_0, α_1, and α_2. The term ϵ_i is the noise in the observations, even if there is no noise in the process of setting the dispenser level, there is general noise in the system. The output after performing linear regression is displayed in table 5.1. The table shows the values of the different parameters corresponding to equation 5.1, their standard deviation and the p-value of the parameter after regression. The parameters have low standard deviation and the p-value is still close to zero. The same analysis is done with the oil concentration from the IR-method. The variable oil_i is exchanged with the variable ir_i that represents the oil concentration found with the IR-method. The resulting model is shown in equation 5.2.

$$ir_i = \alpha_0 + \alpha_1 pr_i + \alpha_2 temp_i + \epsilon_i \tag{5.2}$$

Parameters	Value	Sd.	p-value
α_0	-2450.2	171.0	<0.001
α_1	82.2	4.4	<0.001
α_2	-47.1	4.0	<0.001

Table 5.1: Value of the three parameters in equation 5.1, their Standard deviation (Sd.) and the p-value.

The resulting parameters, after doing regression on the model, are shown in table 5.2. The result is similar to the values in table 5.1.

Parameters	Value	Sd.	p-value
α_0	-2362.7	345.1	<0.001
α_1	77.8	7.6	<0.001
α_2	-45.0	5.4	<0.001

Table 5.2: Value of the three parameters in equation 5.2, their standard deviation (Sd.) and the p-value.

The question is now, how good are these estimates?. One way to visualise the fit is by plotting the data versus the fitted values from the equations 5.1 and 5.2. The figures 5.5 and 5.6 display this relation from the equations respectively. These plots include the best fit regression line. The distance a value appears from the best fit regression line indicates how much it missed the value it was intended to fit. The monitored oil concentrations appears in horizontal lined clusters, in the case of figure 5.5 each such horizontal line represents the five oil concentrations that were used as reference value for the oil dispenser. In the case of figure 5.6 a cluster appears only at 0 ppm where it uses the same data as the oil dispenser. The oil concentration of 0 ppm was assumed to always run pure water in the system because the oil dispenser was turned off. This level of concentration was never monitored with the IR-method.

5.5 Estimation including salinity

The last step in finding a formula that can estimate oil concentration is to include the salinity. The only changes from the equations presented in last section, is the additional salinity term. When adding the salinity term we also include more data in estimation of the predictors in the model.

$$oil_i = \alpha_0 + \alpha_1 pr_i + \alpha_2 temp_i + \alpha_3 sal_i + \epsilon_i \tag{5.3}$$

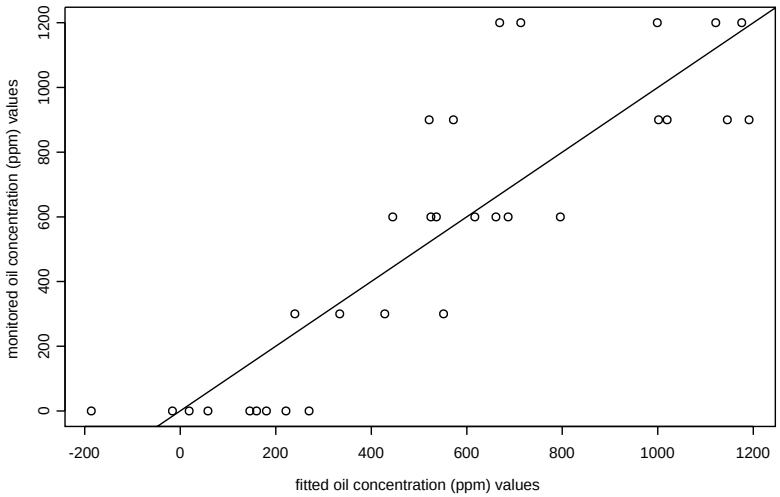

Figure 5.5: Plot of regression fit of the oil concentration from the oil dispenser.

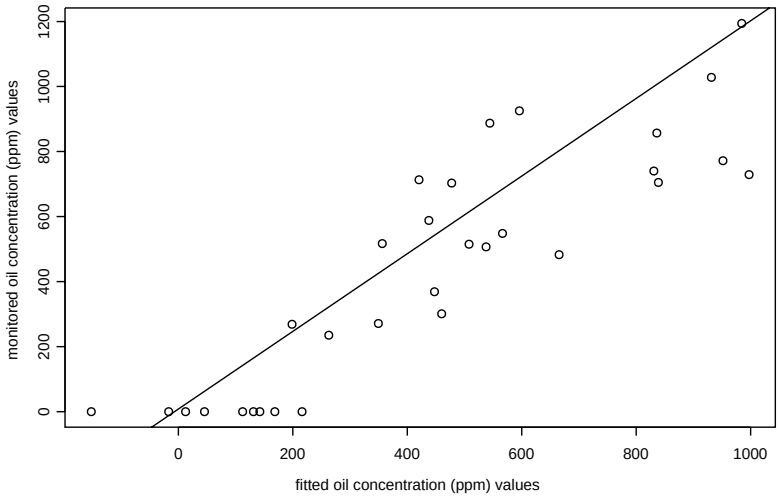

Figure 5.6: Plot of regression fit of the oil concentration from the IR-method.

As before, equation 5.3 uses the oil concentration values from the dispenser to represent oil_i. The only difference from equation 5.1 is the added salinity term sal_i, the other variables have the same interpretation.

Parameters	Value	Sd.	p-value
α_0	-2360.1	179.7	<0.001
α_1	84.3	4.0	<0.001
α_2	-48.4	3.6	<0.001
α_3	-5.2	3.4	0.126

Table 5.3: Value of the four parameters from equation 5.3, their standard deviation (Sd.) and the p-value.

Table 5.3 show estimated parameters and exhibits very similar results as table 5.1, the standard deviation is improved, which should also indicate an improvement of the fit. The p-value of the salinity indicates that the salinity might be unnecessary in the model. This probably occurs because the salinity was tested in the small range of 30-48 g/l. Earlier laboratory experiments with salinity ranges of 0-250 g/l show an increase in photoacoustic response with increasing salinity [3]. The salinity has probably very little influence on the photoacoustic response in the small range tested against. The salinity should not be removed, even if the p-value indicates so, because of the physical dependencies showed in earlier analyses. Now let us do the same analysis on the oil concentration found with the IR-method. The circumstances are the same as in the previous section with regard to the size of the dataset, and the mean of the datapoints. The model in 5.4 is similar to equation 5.3 but the term oil_i is exchanged with ir_i, using the data monitored with the IR-method instead.

$$ir_i = \alpha_0 + \alpha_1 pr_i + \alpha_2 temp_i + \alpha_3 sal_i + \epsilon_i \tag{5.4}$$

The resulting parameters from linear regression are presented in table 5.4 The table 5.4 displays the parameters of the best fit estimate of the

Parameters	Value	Sd.	p-value
α_0	-2090	377.6	<0.001
α_1	77.9	7.3	<0.001
α_2	-45.1	5.2	<0.001
α_3	-9.1	5.6	0.115

Table 5.4: Value of the four parameters, from equation 5.4 after running regression, their standard deviation (Sd.) and the p-value.

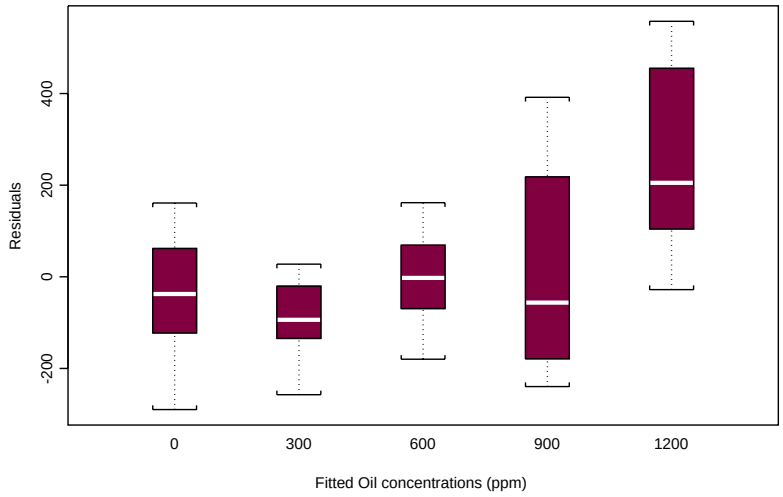

Figure 5.7: Plot of the residual error of the different oil concentration from the oil dispenser, found with equation 5.3.

data, including salinity, but not pressure. As explained earlier no change in photoacoustic response in the tested pressure range was observed. We observe a high p-value with regards to salinity in the table, the salinity should be kept for the same reason mentioned before.

Instead of using the same type of plots displayed in figure 5.5 and 5.6, boxplots of the residuals are used. The definition of the residual error is described in section 4.14. The residual error can be described as the value a fitted observation missed the original observation. In figure 5.7 a boxplot of each of the five oil concentrations from the oil dispenser is displayed. Figure 5.8 displays the residuals of the data monitored with the IR-method. To make the comparing of the two plots easier, the IR-measurements is split into five boxes. Each box spans a range of oil concentration corresponding to $\pm$ 150 ppm around the box. A boxplot contains information on the distribution of the values. We must not forget that we are looking on the residuals, and that the residuals should be centered around zero. The boxplot in figure 5.7 and 5.8 can be interpreted in the following way. The white line in each box gives the median value of the data within that oil concentration range. The black box itself represents the upper and lower quartile which means that the 75% upper limit, and the 25% lower limit, of the distributed values lie within this box. The whiskers is defined as the last value lying within 1,5 $\times$ "interquartile range", where the interquartile range correspond to

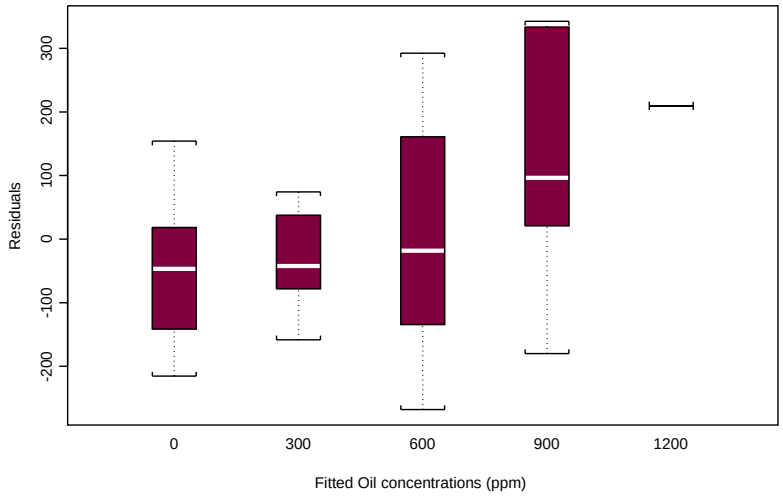

Figure 5.8: Plot of the residual error of the different oil concentration from the IR-method, found with equation 5.4.

the distance between the upper and lower quartile. Any additional lines outside the whiskers describe points that may be classified as outliers.

5.6 Salinity Estimation

As mentioned in section 3.3.2, it might be possible to use the time delay to the first peak in the photoacoustic signal to find the salinity. This time interval will correspond to the speed of sound in water, and is affected by the physical values salinity, temperature and pressure. Already a good deal of work is done on finding equations to express the velocity of sound in seawater. These equations deal with values of salinity, temperature and pressure that are encountered in seawater. The variables in these equations are limited to work within certain ranges. No equations deal with temperatures above $35°C$ or salinity higher than 45 grams per liter (g/l), it is these limitations that make them unsuitable to use as a way to find salinity from produced water.

Water from oil reservoirs often have temperatures above $200°C$ because of the close distance to the earth's core. Produced water then cools down to $50 - 100°C$, before being discharged to sea. The same problem arises with salinity, values from 35 g/l to saturations up to 360 g/l can be encountered in oil reservoirs. With basis in equation 4.25 a model to

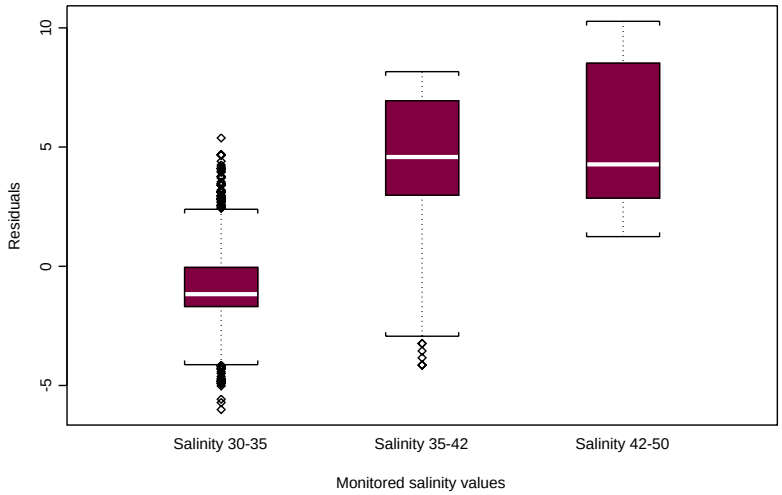

Figure 5.9: Plot of fitted versus the monitored salinity values, found with equation 5.5.

express salinity was created. The pressure variable was omitted because of too little data. After using the trial and error method we ended up with the final model in equation 5.5.

$$sal_i = \alpha_0 + \alpha_1 \frac{1}{temp_i} + \alpha_2 \frac{temp_i}{delay_i} + \alpha_3 \frac{1}{delay_i} \tag{5.5}$$

The variable we want to fit the model to sal_i, the temperature is $temp_i$ and the delay to the first peak is called $delay_i$, for observation number i. Regression statistics are used to estimate the parameters α_0, α_1, and α_3. The estimated values are plotted against the monitored values in figure 5.9, and give an indication on the quality of the model. Figure 5.9 show a boxplot of the salinity estimate. The black boxes is the confidence interval with the mean marked as a white line. The diamonds in the plot indicate outliers, the salinity 30-35 column has a high density of data inside the confidence interval, but still a wide spread that results in many outliers. The salinity of 48 is displayed together with salinity of 45 because of lack of data. From the plot in figure 5.9 the OIWM is able to estimate the salinity with an error of ± 10 in parts per thousand. The model is tested in the range from 30 to 45 g/l, these are not good figures.

Urick [22] states that the salinity has only a small effect on the velocity in seawater, if this is the case it can be hard to get a usable value.

The error in the instrument performance is probably higher than the precision needed to find the salinity. More insight is needed to get a clearer picture on the instruments performance on finding the salinity. Salinity should be tested in a broader range of values, to research further if it is possible to monitor the salinity value.

5.7 Discussion

What can this analysis tell about the performance of the photoacoustic instrument? From figure 5.1 and figure 5.2 the data have strange performance at the higher oil concentrations, compared to earlier observations in laboratory experiments. Is this the fault of the model, the instrument or an unstable test environment?.

Trying to fit lines to each of the five oil concentrations reveal that the fitted lines do not match the expected behavior at higher oil concentration corresponding to higher response. A few problems are known to exist when doing monitoring. When the instrument is inserted in the pipeline and oily water is passing, a thin film tends to fasten on the instrument window where the laser enters the sample. The oil film results in a heightened photoacoustic response during monitoring. The effect can be viewed as a noise phenomena, and is coloured in nature. The figures 5.7 and 5.8 show that the residual value increases with higher oil concentration. This indicates a systematic error that is probably caused by the oil film. The instrument window was cleaned three times during the Porsgrunn tests, and each time the photoacoustic response went back to the expected photoacoustic response for 0 ppm. This is a problem that must be solved to give a sufficiently accurate estimate of the oil concentration.

The thesis has also shown that the instrument's accuracy in measuring salinity is not sufficient in the monitored range. There is a chance that the instrument is able to detect salinity changes within an error of $\pm$ 10 parts per thousand. This fact must be researched further for a larger range of salinity values.

Another problem during the testing was the monitoring of the reference oil concentration. The oil concentration at each test point is not exact, this gives problems with estimation of the data, the IR-method is known to have an error of 10%. In addition, the performance of the oil dispenser has not been properly tested. It is not known if it has a linear output for increasing oil concentrations. The IR-method should detect this, but there are few monitored values for higher oil concentrations. In the cases where two IR-samples have been collected at the same time at high oil concentrations they differ greatly. An analysis of the oil dispenser's stability during a test sequence has not been performed. It will

be reflected in the data if the oil concentration of the dispenser differs during a run.

By running the data through equation 5.4 and looking at the output, some observations can be made. The data within a test sequence (time interval with constant salinity, temperature and oil concentration) is coherent and the difference in estimated value is usually around 100 ppm, the worst unstable cases have a difference around 200 ppm. Often the mean in different datapoints is not coherent even if the basic test parameters are the same, some of the reasons that might influence this problem are mentioned above.

Three cases should be considered, the test rig does not give a correct concentration, an oil layer on the instrument's window builds up over time and causes an unstable reading and/or the instrument is drifting. In the first case the solution is to keep better track of the oil concentration to ensure that the rig gives a stable and repeatable concentration. Build up of oil on the instrument window was observed during the tests in Porsgrunn. The photoacoustic signal corresponding to 0 ppm did go down when the instrument was retrieved from the pipeline and cleaned. Drift in instrument reading can be lack of warming-up of the instrument before testing, or the use of unsuitable electronics. Prior to the Porsgrunn test the instrument went through a stability test of 1-2 weeks. It was then tested towards water and no drift was observed over time. Hence, as long as sufficient warm up time of the instrument is provided, there should not be any drift in the readings if the instrument is exposed to the same input. In this case the problem is to keep better track of the oil concentration to show stability. Hopefully case one is the problem, together with erasing the oil film problem, it should be possible to make estimates with an accuracy of $\pm$ 100 ppm. If combined with some kind of signal processing algorithm, even better results can be expected.

Chapter 6

Analysis of the OIWM signal

6.1 Introduction

The OIWM has so far depended on the peak-to-peak value to find the oil concentration. Little work have been performed on finding more efficient methods to analyse the signal. Some work on Fourier and wavelets have been done [5], trying to gain frequency information in time that can find patterns in the signal and retrieve new information about the sample. With a filter the noise in the signal can be removed and the instrument gain better accuracy and stability. In this chapter some filters will be explored to show the general effect on the signal and how this affects the peak-to-peak value. Later spectral moments will be used as an alternative method to the peak-to-peak value.

6.2 Filter specifications

Filtering is helpful when there is a need to remove certain frequency components. In the OIWM signal a lot of high frequency background noise exists. Low frequency noise from pumps and valves can appear in the production environment, filtering will greatly reduce such disturbance. The pre-amplifier already does some filtering, but no data exists on the general effect on the signal. A typical signal is illustrated in figure 3.4. Taking the Fast Fourier Transform (FFT) of this signal result in figure 6.1. The figure shows one tenth of the spectrum, the magnitude is scaled to dB for illustrative effect. The spectrum shows that very little frequency information exists beyond 2 MHz. The strongest components in the signal are in the band from 0,1-0,6 MHz, so a passband within these ranges is wanted. The influence of the spectral components lower than 0,1 MHz is not known, but after working on the signal for some time it is decided to filter the range. Filtering of the low frequencies will also remove possible noise from pumps and valves. A sharp transition

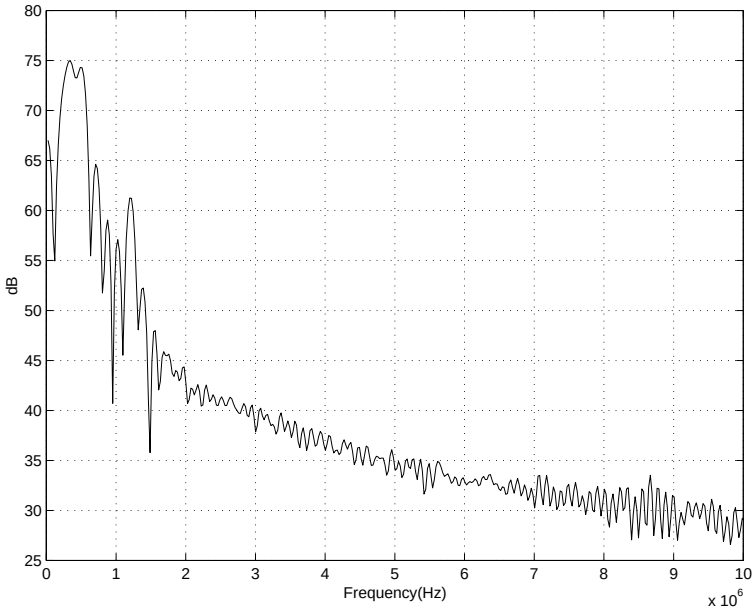

Figure 6.1: Plot of the Fourier transform of a normal OIWM signal.

band is wanted, but the final specifications depends on the filter construction. The stopband should be at least -20dB, but going lower is only an advantage. Two types of filters were tried, a Finite Impulse Response (FIR) filter and an Infinite Impulse Response (IIR) filter. The exact shape of the filters was achieved by trial and error using Matlab [26]. The Matlab code used to construct the filters can be found in appendix A. The design of filters is well documented [20, 19], and there are many application to help in the design.

6.2.1 FIR filter

First a FIR filter was constructed. The filter is created with the window method using a Hamming window. The order of the filter is 500. The high order was judged necessary for the filter to cut at the lowest frequencies, and at the same time have a relatively sharp transition band. The magnitude and phase of the filter is displayed in figure 6.2. The characteristics of the transition band could be sharper, but that would include raising the order. When filtering over such a small passband it is nearly impossible to get sharp edges. The stopband is rippled, and laying at -40dB which gives good filtering. Other windows (kaiser,triangular) were tried and gave similar results with only small changes in filter properties. Another advantage of the FIR filter is that it

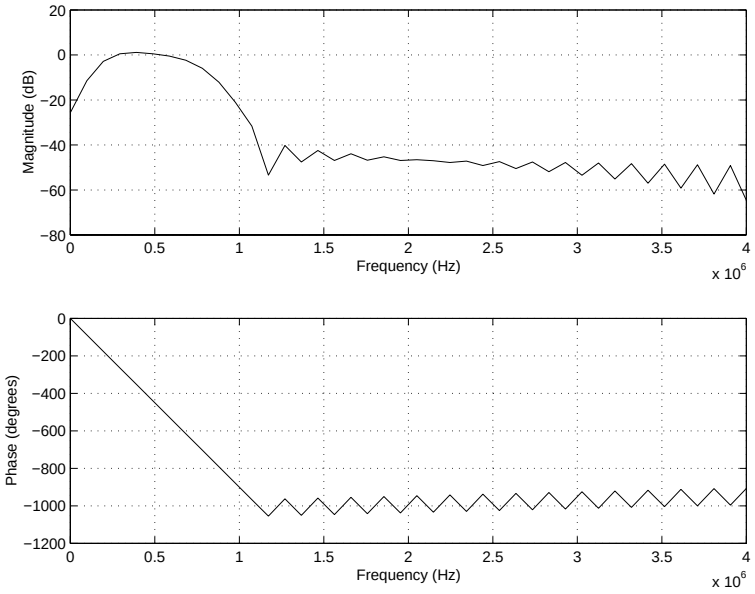

Figure 6.2: Order 500 FIR filter with linear phase.

retains linear phase, thus all frequencies undergo the same time delay. In contrast to an IIR filter a FIR filter is almost entirely restricted to discrete time implementations.

6.2.2 IIR filter

Simulation of an IIR filter involves constructing a continuous time filter and then transforming it into a discrete time filter. In this case a Chebyshev type II bandpass filter with order three is used. The magnitude and phase of this filter are shown in figure 6.3 The stopband is at -20 dB and gives fairly good filtering. The filtering is also good at the low frequencies. The Chebyshew II filter has a flat stopband and a rippled passband. The filter displays a nonlinear phase, something that results in a dispersion in time delay of the filtered signal.

6.2.3 Comparison of filters

The original signal is plotted together with the IIR filtered and the FIR filtered signal in figure 6.4. The original signal and the FIR filtered signal follow approximately the same path. The FIR filter has removed the high frequency noise, and kept the most important components in the signal. The peaks of the FIR signal have the same placement, and little change in amplitude occurs. In comparison the IIR filtered signal also removes

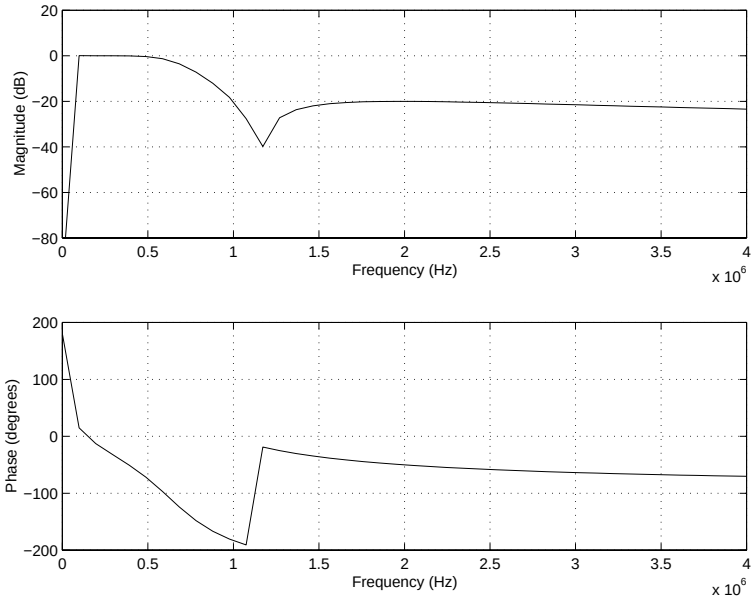

Figure 6.3: Chebyshev *II* IIR filter of Order 3 with non linear phase.

the high frequency noise in the signal. It is displaced in time, this is not a big problem when we are looking at the peak-to-peak value. The problem is the change in amplitude of the peaks, this phenomenon occurs because of the non-linear phase of the filter. Some of these effects may be avoided if we had used a Bessel type filter (maximally flat time delay) instead, but software to compute such a filter was not available. The peak displacement can result in unwanted effects when finding the peak-to-peak value. Very careful design is needed if an IIR bandpass filter is constructed that will not alter the peak-to-peak value of the desired signal.

6.2.4 FIR filtering of the dataset

Improvement as result of using IIR filter was not visible, so only the result from the FIR filtering is shown. The photoacoustic signal was convolved with the FIR filter described in section 6.2.1, and the peak-to-peak value was then retrieved from the filtered signal. The output from this process was then plotted against temperature and the result is displayed in figure 6.5. The figure produced is similar to some of the plots in chapter 5. Although the plots are not comparable because they use two different datasets. The waveform data used in this section is averaged over fewer laser pulses than the dataset used in chapter 5.

The plot in figure 6.5 uses the oil concentration from the oil dis-

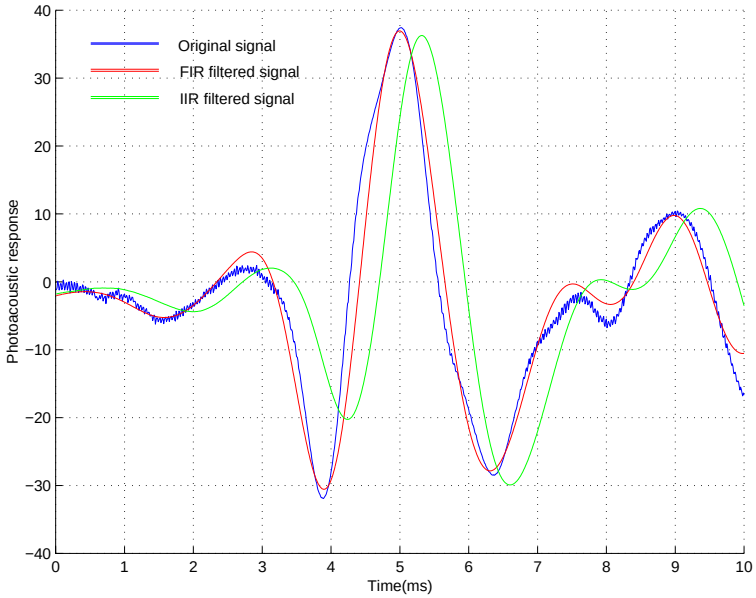

Figure 6.4: Plot of original, IIR filtered and FIR filtered signal.

penser as reference. It displays the temperature versus the peak-to-peak value of the filtered signal, with no regard to salinity and pressure. The oil concentration is colour coded to make a distinction between the different levels. The plot of the unfiltered peak-to-peak value is also plotted against temperature and colour coded in the same way. This plot can be used as a reference for comparison, and is displayed in figure 6.6. The two plots are similar in appearance, both methods give a reasonable presentation of the oil concentration. Which is best should be decided with some kind of statistical analysis, this is not done in this thesis because the statistics concentrate on evaluation of the spectral moments. If filtering of the signal on average gives the same effect as the non-filtered, a few benefits should be noted. The filter will attenuate unwanted noise that exists in a production environment. Such noise can be coloured and appear at uneven time intervals. Such disturbance can be noise created by valves, vibrations in the pipe line, or interference from nearby electronic instruments. The test environment was not accommodated to such conditions.

6.3 Spectral Moments used in signal determination

A filter will try to filter out the unwanted frequencies, and deliver a new version of the signal. Because of this procedure a filter has tendencies to

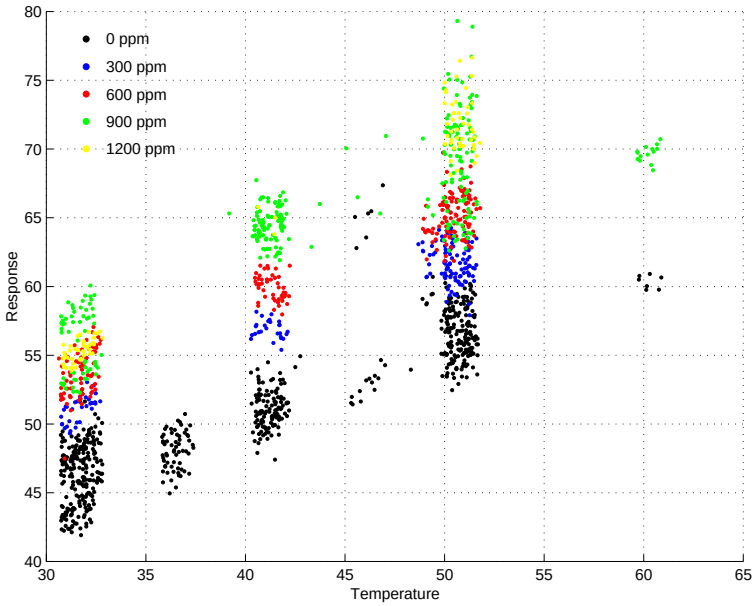

Figure 6.5: Relation between the peak-to-peak value of the FIR filtered signal and the temperature.

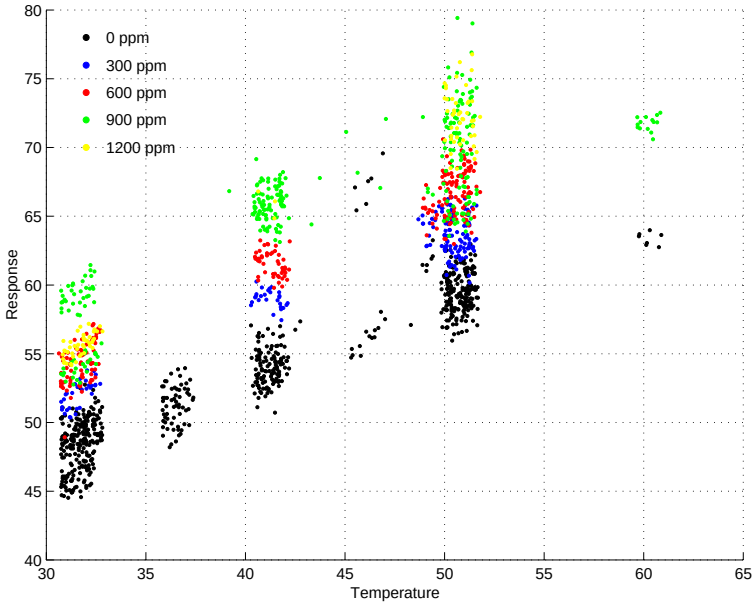

Figure 6.6: Relation between the peak-to-peak value plotted against temperature.

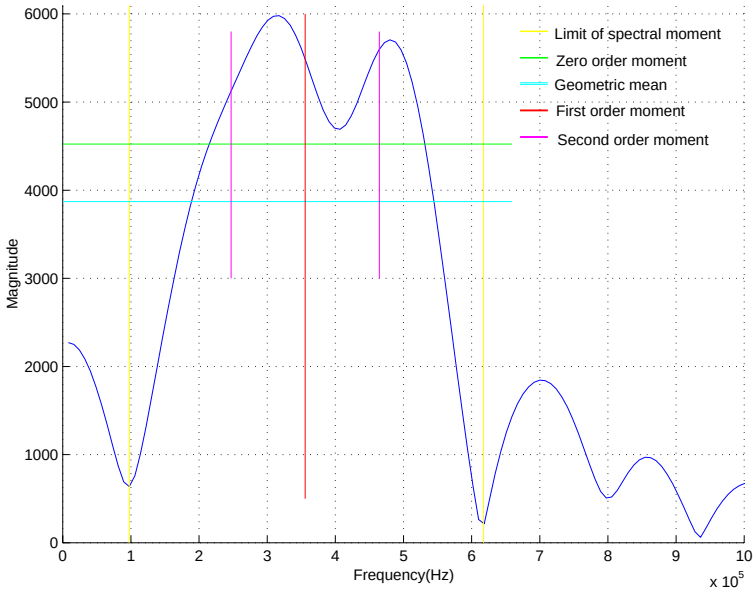

Figure 6.7: Visualization of the three first spectral moments.

distort the shape of the signal, as is the case of non-linear phase filters. Because little was gained in accuracy when filtering, alternative methods to represent the photoacoustic response were seeked.

One such method is to use the spectral moments, the theory is described in section 4.1. Spectral moments are usually equated over the whole spectrum. In this case we want to remove the influence of frequencies that can contribute to unwanted noise in the signal. The frequencies between 0.1 MHz and 0.62 MHz are the strongest frequency components in the signal, and were picked as appropriate to use in finding of the spectral moments. These limits are marked off with yellow lines in figure 6.7. The figure is a zoomed in Fast Fourier Transform (FFT) of the photoacoustic signal. For illustrative purposes the magnitude of the spectrum was used instead of the power spectrum. The first three moments and their interpretation is illustrated in figure 6.7. Each moment is colour coded and described below. The zero order moment is defined as the mean power of the signal, and is found with equation 6.1.

$$\overline{m}_a = \frac{1}{N} \sum_{l=C_1}^{C_2} S(l) \tag{6.1}$$

Another value that gives a similar representation to the zero order moment is the geometric mean. This value is related to the arithmetic mean,

and interesting because it is easy to equate.

$$\overline{m}_g = \Big(\prod_{l=C_1}^{C_2} S(l) \Big)^{1/N} \tag{6.2}$$

The two methods have similar results, but the geometric mean favors the high values in the spectrum. The arithmetic mean is displayed in green in figure 6.7, and the geometric mean is displayed with cyan in the same figure. The geometric mean came to the attention when the Fourier transform was displayed logarithmically. This resulted in the use of the logarithm on each element before equating the mean magnitude of the spectral moments. The result was interesting so the case was further researched. The relation between the geometric mean and the arithmetic mean of the logarithmic spectrum is shown in equation 6.3.

$$\begin{aligned} \tfrac{1}{N}\textstyle\sum_{n=1}^{N} 20log x_n &= 10\textstyle\sum_{n=1}^{N} log(x_n^2)^{1/N} \\ &= 10(log(x_1^2)^{1/N} + log(x_2^2)^{1/N} + \cdots + log(x_n^2)^{1/N}) \\ &= 10log(\textstyle\prod_{n=1}^{N} x_n^2)^{1/N} \end{aligned} \tag{6.3}$$

The mean used on the logarithm of the power spectrum, gives a scaled version of the geometric mean, scaled by 10 times the logarithm. When finding the oil concentration we are after a proportional number so the change in scale will not disturb the result. Let us return to figure 6.7 and look on the first order moment. This is interpreted as the mean frequency and is displayed with a red line. The second order moment is the standard deviation of an equal Gaussian spectrum. It is a measure of the width of the signal about the axis of the first order moment. The second order moment is displayed in magenta and equals the distance between the two lines. Higher order moments are impossible to display graphically in two dimensions. The third order moment is a measure of the skewness in the spectrum, this value will be zero if the spectrum is totally symmetrical about the zero order moment.

6.3.1 Using spectral moments to find oil concentration

Instead of using the peak-to-peak value we can use the output from the spectral moments. Each moment has its own interpretation as described above, and might distill information that can improve the result when finding the oil concentration. To illustrate that the spectral moments are a good alternative method to the peak-to-peak value, a plot of the zero order spectral moment run on the data is displayed. Of the spectral moments the zero order moment has the largest similarity to the peak-to-peak value. The other moments may still contain information that can help in determining the oil concentration. It is a possibility that

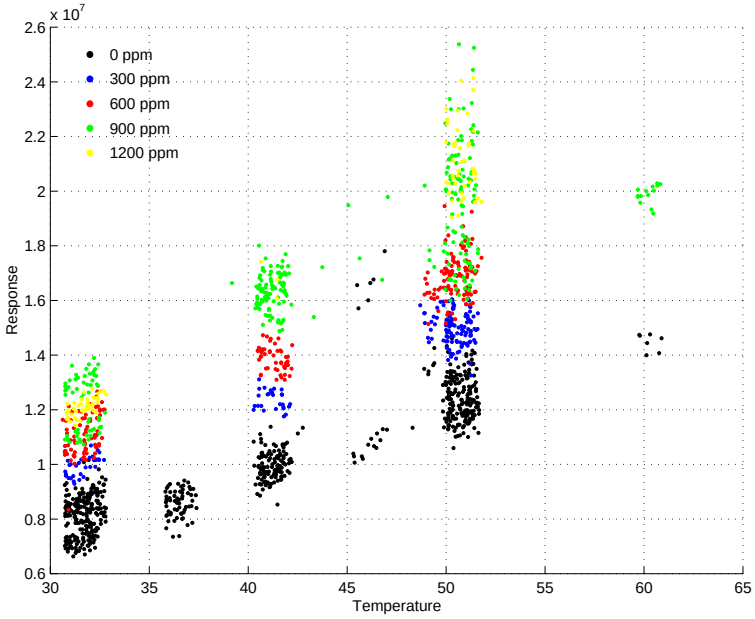

Figure 6.8: Plot of arithmetic mean.

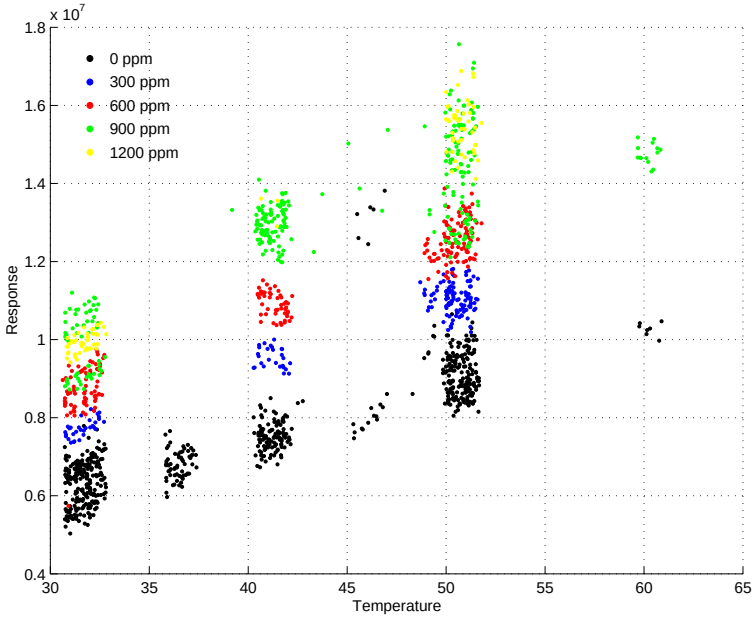

Figure 6.9: Plot of geometric mean.

this connection can be detected in a statistical analysis, and will be discussed in the next chapter. Both the output from the arithmetic mean (zero order moment), in figure 6.8, and from the geometric mean, in figure 6.9, are displayed. Both figures show good clustering between the different oil concentrations. The geometric mean appears to have better separated clusters, this is probably because the high values in the spectrum contains more information about the oil concentration, and the geometric mean favors these values.

6.4 Finding m0 in the time domain

When looking for an instrument that can be used in a downhole installation it is important to find methods that make the data aqusition more effective. This section proposes a method that will reduce the need of computer software when equating the photoacoustic response. The computation is reduced to a set of filtering, the filters are discrete, but it should be possible to transform the process to the continuous case. In section 6.2 a FIR bandpass filter and an IIR bandpass filter were proposed as candidates to filter the signal. The FIR filter was the best candidate because it retained a linear phase. The high sampling frequency makes it difficult to produce an efficient bandpass filter, so the FIR filter proposed earlier has an order of 500, this is high and not recommended for use in future implementations. This is the reason the filtering process is split in two stages this time. First the signal is filtered with a lowpass filter, this attenuates the high frequency components in the signal and helps to avoid aliasing when sampling the signal at a lower frequency. Next step is to create a highpass filter that removes the low frequency components below 0.1 MHz. This is hard to achieve with the high sampling frequency of 100 MHz. Thus the signal was first decimated before the highpass filter was applied. The decimation process corresponds to sampling at a lower frequency, and shows that it is not necessary to use the high sampling frequency of 100 MHz used today. Even with a lower sampling frequency it will be fairly easy to reconstruct the lowpass filtered signal by zero padding and interpolation.

6.4.1 Applying the lowpass filter

The lowpass filter is an FIR filter using a Kaiser window with order 50 and β equal to 0,5. These specifications give a high sidelobe level, traded with a narrow mainlobe. The filtering in the stopband is $-20dB$, beginning at about 2,1 MHz. The magnitude and phase spectrum of the filter is shown in figure 6.10. The order of the filter is reduced to 50, that is 1/10th the order of the filter used in section 6.2.1. This gives a filter

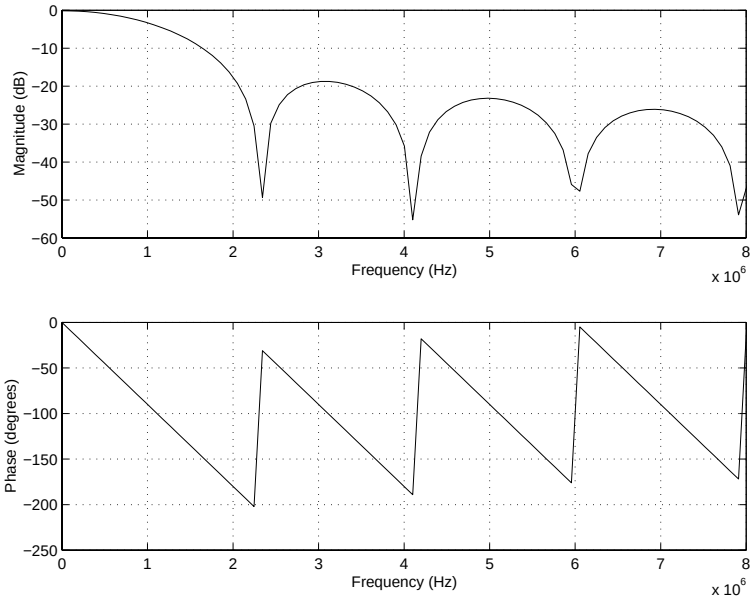

Figure 6.10: Properties of the lowpass filter.

with poorer characteristics, but much more realistic to implement both in hardware and software. It is still important to keep a linear phase to avoid distortion of the peaks in the signal. If using a Bessel type of filter it should also be possible to keep these specification with a IIR filter. The signal was then convolved with the filter to get a filtered signal that can be processed further. If the lowpass filter was implemented in hardware the next step would be to sample the signal with an A/D converter. The signal is already sampled so we decimate the signal to get the corresponding effect as sampling at a lower frequency.

6.4.2 Decimation and highpass filtering

Before applying the highpass filter the signal was decimated. The signal was decimated down to 1/25th of the original size, of 1000 samples only 40 were kept. This corresponds to a sampling frequency of 4 MHz. It should be possible to sample at an even lower frequency, but this must be further researched [21]. After decimation the highpass filter was applied. The filter is a simple free-hand construction. It is a linear phase FIR highpass filter based on a moving average filter [19]. The length equals the length of the signal, and the impulse response, $h[n]$,

of a length M filter is expressed in equation 6.4.

$$h[n] = \begin{cases} 1 - \frac{1}{M+1} & n = \frac{M+1}{2} \\ -\frac{1}{M+1} & 0 < n < M+1 \\ 0 & otherwise \end{cases} \tag{6.4}$$

The impulse response h[n] is plotted in figure 6.11, the filter is causal and the symmetry about $\frac{M+1}{2}$ gives the filter linear phase. When con-

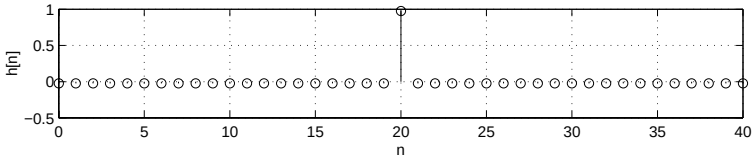

Figure 6.11: The impulse response of the highpass filter.

volved with the signal the filter will pass high frequency components present in $n = \frac{M+1}{2}$, and attenuate the effect of more slowly varying frequencies. To better understand the properties of this filter the magnitude and phase plot are displayed in figure 6.12. The next step is to convolve the signal with the highpass filter. The output is displayed in figure 6.13. The filtered signal is displayed together with a decimated version of the original signal for comparison. The original decimated signal is in red, and the filtered signal in blue. From the filtered signal it is important to find a good representation for the photoacoustic response. There are three ways to find this value. The first method is the peak-to-peak value of the signal, this will correspond to the procedure performed in section 6.2 and gives a similar result. The second method is to use the arithmetic mean power. The third method is to make a summation over the magnitude of the signal. The arithmetic mean power of the signal is related to the frequency domain by Parseval's theorem [20]. Parseval's theorem is valid for the power spectrum of Fourier transformed signals. In the discrete case the relation is expressed in equation 6.5

$$\sum_{n=0}^{N-1} |x[n]|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |X[k]|^2 \tag{6.5}$$

Where x[n] is an N-point sequence, and X[k] is the N point FFT of the signal. From equation 4.5 we have the relation for m_0 reproduced in equation 6.6.

$$m_0 = \frac{1}{I} \sum_{l=c_1}^{c_2} S(l) \tag{6.6}$$

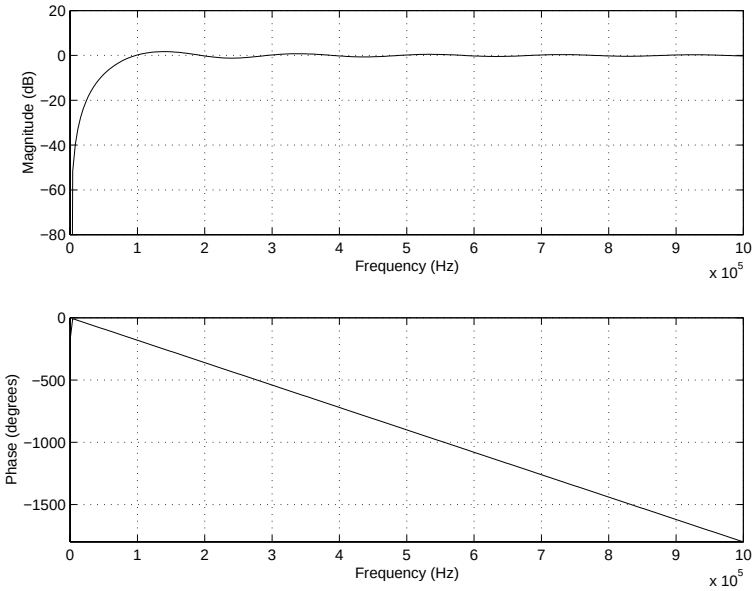

Figure 6.12: Properties of the highpass filter.

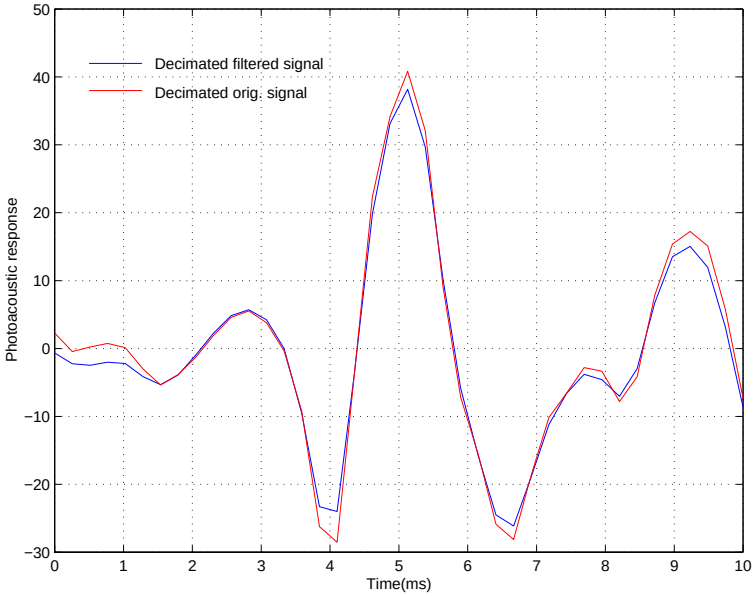

Figure 6.13: The decimated signal, the filtered in blue and the original in red.

Where $S(l)$ is the discrete spectrum of length $I = c_2 - c_1$. The term $S(l)$ equals the term for the power spectrum $|X[k]|^2$ from equation 6.5. Equation 6.7 gives the following relation in the time domain for m_0.

$$m_0 = \sum_{n=0}^{N-1} |x[n]|^2 \tag{6.7}$$

This shows that it is possible to equate the zero order spectral moment directly from the time domain itself. If filtering out unwanted frequencies first, we get the same effect as using the zero order moment over a range of frequencies in the spectrum. This method performed well. The scale is different, this is because the length of the FFT is different from the length of the signal. Running the dataset through this method produced a similar results as the zero order spectral moment. The plot of this result is displayed in figure 6.14. When comparing figure 6.8 with figure 6.14 the two plots are indiscernible. This shows that the total energy in the signal can be used to find the oil concentration. The decimation process is important to make the computation of the energy fast and efficient. Instead of depending only on the two values used by the the peak-to-peak method, this method works on the whole signal. At the same time the method is easy to implement both in hardware and software. Another and by far the simplest method to find the photoacoustic response is by averaging the a filtered signal. In this case we wish to do ordinary averaging with equation 6.8.

$$m_0 = \frac{1}{N} \sum_{n=0}^{N-1} |x[n]| \tag{6.8}$$

In a hardware implementation a standard averaging filter can be used to outputs the photoacoustic response value at certain time intervals. This value is not related directly to the spectral moments in the same way as the signal energy in equation 6.7. It is inspected closer because of the ease of implementation and close similarity to the signal energy. The average of the filtered signal and the statistical performance is compared to the other techniques in the next section. This can be a possible implementation when running the instrument in a downhole environment, because it removes the need of a computer to equate the photoacoustic response.

6.4.3 A statistical view of the data

It is interesting to take a look at the statistical similarity between the parameters of the zero order moment, the signal energy and the average of the signal. A mathematical connection exists through Parseval's

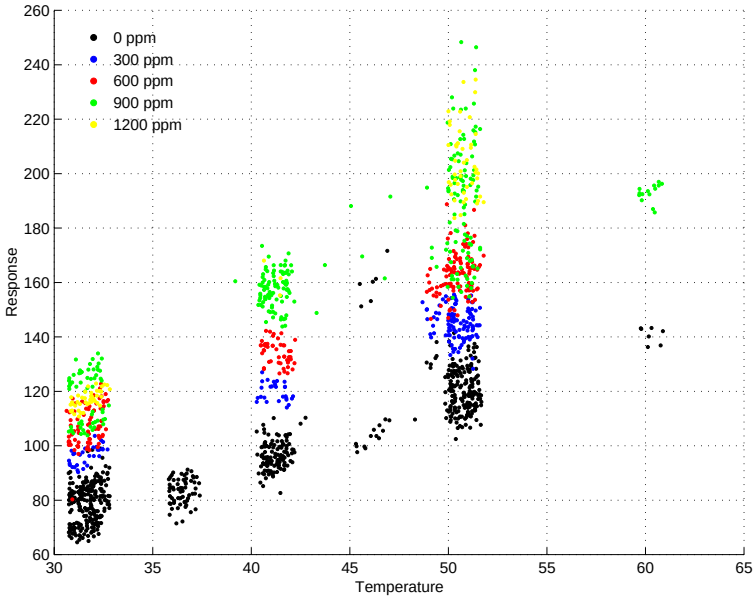

Figure 6.14: Output from using the arithmetic mean on the decimated and filtered signal.

theorem between the zero order moment and signal energy, statistics shows that the signal average give similar performance. It is hard to get total equality in the output from the three methods. This is because a couple of Parseval's criteria are broken. Parseval's expects that the N-point sequence is transformed to an N-point Fourier sequence. Parseval's also expects that the whole spectrum is used not only parts of it as done when equating the zero order moment. To show partial equality we wish to make an estimate with linear regression, and compare the outputs. The models used are presented in equation 6.9, 6.10, and 6.11

$$oc_i = \alpha_0 + \alpha_1 m0_i + \alpha_2 temp_i + \alpha_3 sal_i + \epsilon_i \tag{6.9}$$

$$oc_i = \alpha_0 + \alpha_1 filter_i + \alpha_2 temp_i + \alpha_3 sal_i + \epsilon_i \tag{6.10}$$

$$oc_i = \alpha_0 + \alpha_1 average_i + \alpha_2 temp_i + \alpha_3 sal_i + \epsilon_i \tag{6.11}$$

Here oc_i is the oil concentration using the values from the oil dispenser, where the subscript i represents observation number i. The term $m0_i$ in equation 6.9 is the zero order moment. The $filter_i$ uses the mean energy of the filtered signal, and $average_i$ use the average of the signal. The $temp_i$ and sal_i represent the temperature and salinity respectively. All the variables were normalized before doing regression, this makes the output parameters unusable for calibration of the instrument, but

suitable for comparison of equations. The output parameters are displayed in table 6.1. The table display values that are close to each other.

Parameter	Equation 6.9	Equation 6.10	Equation 6.11
α_0	-5.0337×10^{-16}	-2.6000×10^{-16}	$3.5964 \times 10^{-}16$
α_1	0.7753	0.7736	0.8313
α_2	-0.2418	-0.2428	-0.2092
α_3	-0.1821	-0.1816	-0.1329

Table 6.1: The table display the value of the four parameters in the three equations 6.9, 6.10 and 6.11 after running linear regression.

They are not equal, but indicate that there a connection exists. The residual sum of squares (RSS) is 601 for equation 6.9, 606 for equation 6.10 and 445 for equation 6.11. Surprisingly the signal average receive the best result. A result that will be studied more closely in the next chapter. This should indicate that the performance is almost identical for equations 6.9 and 6.10. The result when using the ir-measurements from the oil dispenser have a similar appearance to the numbers presented above. A bonus is that the signal average are easiest to implement and at the same time gives best result.

6.5 Discussion

This chapter proposes some alternative methods to the peak-to-peak value. First the filtering process is important to reduce both high and low frequency noise. The components bearing information in the signal are found to be in the range 0,1-0,6 MHz, so passing these components should be an important characteristics of the filter.

The spectral moments are a totally new way to analyse the signal, they work directly on the spectrum and can be limited within these ranges. The zero order moment is the best candidate to replace the peak-to-peak value. Little is known of the influence of the higher order moments, but there might be important information that can be used to get a better representation of the oil concentration. Statistics will be used in the next chapter to analyse the effect of the higher spectral moments.

The analysis also shows that the sampling frequency can be greatly reduced. Here it is important to have a lowpass filter that smooths the signal, and removes possible aliasing when sampling. The sampling can be safely reduced down to 1/25th of the original signal without loss of peak-to-peak information. It should be possible to sample the signal at even lower frequencies. When using low sampling an alternative method

building on the zero order spectral moment is proposed. Instead of using the peak-to-peak value it is possible to use the arithmetic mean power over the signal to represent the oil concentration. When using this method it is necessary to perform some filtering of the signal, but it is probably sufficient to use a lowpass filter before sampling. In experiments with this method the signal was decimated down to 1/100th of the original signal, this represent a sampling frequency of 1 MHz. This process still shows good ability to represent the photoacoustic response. The parameters from linear regression illustrate that both the arithmetic mean power and average of the filtered signal produce similar results to the zero order spectral moments. This makes it possible to make a fast and stable method for monitoring. The photoacoustic response can then be produced with the work of a highpass, lowpass and an average filter. If the hardware filters is constructed with care, the A/D card and computer can be removed in future applications.

Chapter 7

Model selection

7.1 Statistical methods

An alternative method to the peak-to-peak value when finding the oil concentration is to use the spectral moments previously described. Both methods display values that look sensible. In this chapter statistical methods are used to select between different variables and models, trying to find the model with the best properties. If one model possess better qualities, some considerations must be taken. It is important to evaluate the behavior towards background noise and the complexity of calculation.

This chapter will try different statistical tools to compare the performance of the different methods analysed in this thesis. S-plus was the tool of choice for the analysis. Because of big differences in the value of the spectral moments and peak-to-peak value, the data was normalized before using them in the analysis. This avoids possible problems with scaling, and makes it easier to compare parameter values and standard deviation. The output will however be wrongly scaled and only usable in statistical analyses. We use the same dataset as in the previous chapter concerning analysis of the spectral moments. The peak-to-peak value is derived from the wave signal, and not the more accurate peak-to-peak value produced by the instrument. The variables used in the analysis are the three first spectral moments, geometric mean, peak-to-peak value, temperature and salinity modeled to fit the oil concentration expressed in ppm. To get an acceptable size on the datasets with regards to estimation of salinity, only the oil concentration from the oil dispenser is used. Instead of the zero order spectral moment the geometric mean was used because it has better ability to model the data. The zero order moment is brought into the analysis in the section 7.4.1 that treat *Cp*-criterion on multiple models.

	Value	Sd.	t-value	p-value
α_0	4361.2608	223.2275	19.5373	<0.0001
α_1	919.1997	25.0836	36.6454	<0.0001
α_2	52.5483	8.0643	6.5161	<0.0001
α_3	110.7097	7.2987	15.1685	<0.0001
α_4	-66.0734	3.7504	-17.6176	<0.0001
α_5	-18.5297	4.1829	-4.4299	<0.0001
α_6	-7.6656	4.1421	-1.8507	0.0644

Table 7.1: Output from regression of equation 7.1. The parametric values are displayed with the standard deviation (Sd.), t-value and p-value.

7.2 The p-test

The p-test is a method widely in use to evaluate parameters influence on a model. If one of the parameters output a sufficient low value from the p-test, the parameter is counted as significant in the model. In chapter 4.3 the theory behind the p-test and the t-test are described. We want to find out which of the variables in question are significant. Both the peak-to-peak value and the geometric mean, first and second order spectral moments together with salinity and temperature are included in the model. A fit to the model is made with linear regression, and the output is then compared with a p-test. The model is presented in equation 7.1.

$$\begin{aligned} oc_i &= \alpha_0 + \alpha_1 geo_i + \alpha_2 m1_i + \alpha_3 m2_i \\ &\quad + \alpha_4 pktopk_i + \alpha_5 temp_i + \alpha_6 sal_i + \epsilon_i \end{aligned} \tag{7.1}$$

In equation 7.1 the oc_i represents the monitored oil concentration, in this case the values from the oil dispenser are used. The term geo_i is the geometrical mean of the Fourier spectrum ,$m1_i$ and $m2_i$ are the first and second order spectral moments. The variables $pktopk_i$, $temp_i$ and sal_i represent in the following order; the photoacoustic response found with the peak-to-peak value, temperature and salinity. The noise in the model is ϵ_i and is assumed to be normal distributed with zero mean and variance of σ^2. Each variable is valid for observation number i.

Using linear regression we get the parameters in table 7.1. The last column is the p-value. The p-value of each parameter is close to zero, except the value for salinity. This indicates that the salinity is the only predictor that appears to be insignificant. Previous testing have shown that the photoacoustic response is affected by salinity, and there is no reason to remove it from the analysis. The conclusion is that all variables in equation 7.1 have some influence on the fit of data. The purpose of this chapter was to decide between the peak-to-peak value and the spectral moments, so there are a need for more statistical analysis.

	Sum of Sq	RSS	*Cp*
current		38916842	39255249
α_1	32460133	71376975	71667038
α_2	1026338	39943180	40233244
α_3	5561548	44478390	44768453
α_4	7502447	46419289	46709353
α_5	82787	38999629	39289693
α_6	474343	39391185	39681249

Table 7.2: *Cp*-values after stepwise removal of variables in model 7.1. The Residual Sum of Squares (RSS) and the *Cp*-value is displayed for each case.

By looking at the correlation plot between the different covariates we get an indication of how the variables relate to each other. Such a plot of can be seen in figure 7.1 on the facing page. The diagonal contains the different covariates from equation 7.1, where the square in the upper left corner marked ppm is the reference oil concentration, geo is the geometric mean. The squares m1 and m2 represent the first and second order spectral moments, the square marked sal is salinity and temp equals the temperature covariate. To look at the correlation between two variables cross reference the lines where the squares are present. The plot indicates that zero order moment and the peak-to-peak value are strongly correlated. Another connection that should be noted is the correlation between the second order moment and temperature.

7.3 *Cp* model selection

The p-test in the previous section gave no good answers of what might be a preferable model, only the significance of a specified variable. In this section we use a model select criterion called *Cp*[1]. Basically *Cp* is a method that is used to compare different models based on the residual sum of squares. The method, and the mathematics behind are described in section 4.3. We use the same model described in equation 7.1. A low Cp-value indicates a good result. In this case one covariate from equation 7.1 is removed, and then the *Cp*-criterion on the resulting model is used. We do this operation for each term in the model. This way the influence of a each variable is explored. If a variable displays a high Cp-value when removed, the variable has strong influence on the model and is significant. The top line in table 7.2 represents the *Cp*-value of the current model, and is expressed in equation 7.1. If the current model

[1]The procedure used in S-plus is called step

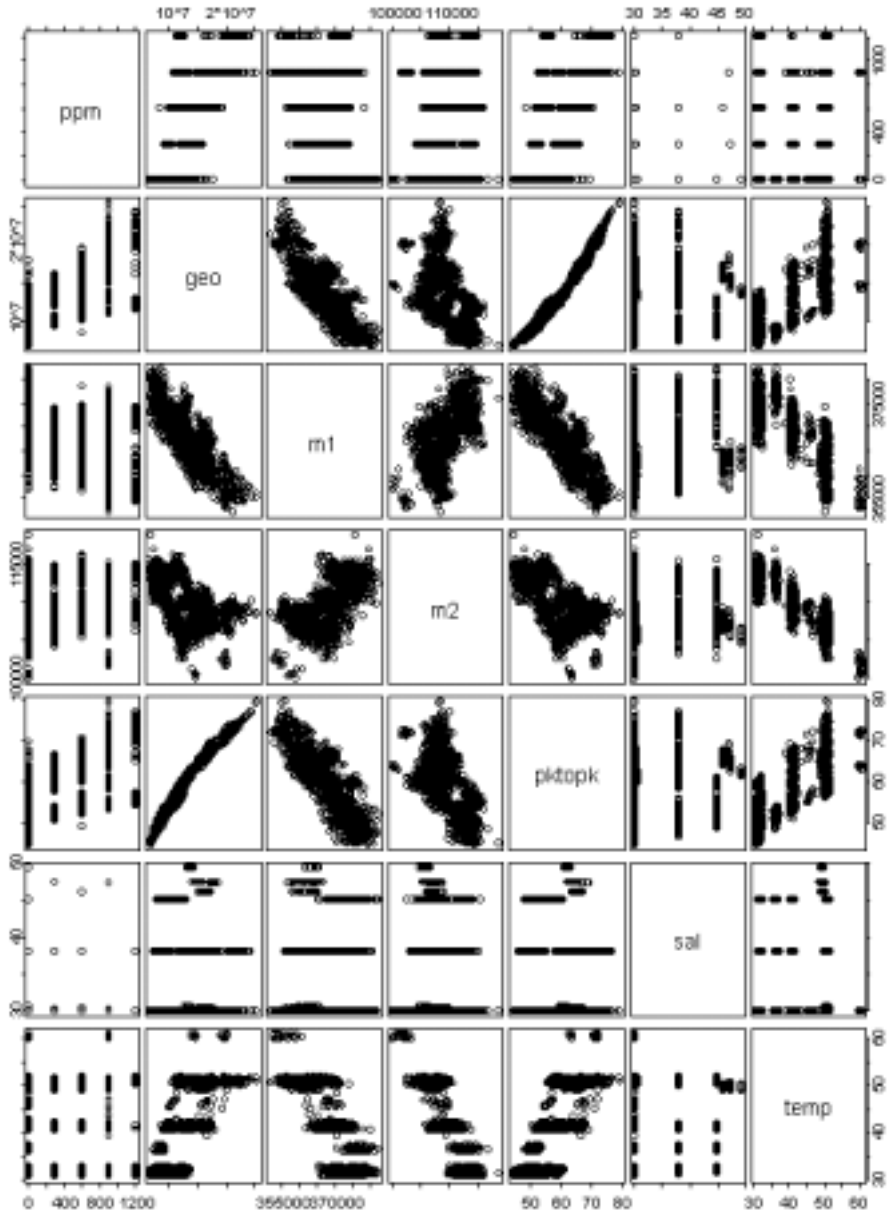

Figure 7.1: The variables from equation 7.1 plotted against each other, showing correlation. The squares along the diagonal contain the variable names. From the top left corner: ppm equals the reference oil concentration, geo, m1 and m2 is the three first spectral moments, sal the salinity, and temp the temperature.

displays the lowest *Cp*-value, the most advantageous model assessed by the *Cp*-criterion, is to keep all terms. From table 7.2 the α_1 has a higher value than α_4, this tells us that removing α_1 from the current model gives a worse *Cp*-value than removing α_4. This indicates that the zero order spectral moment has larger influence on the model than the peak-to-peak value. Because the original model receives the best Cp-value, no variables can be judged unimportant as a result of this test. We want to find a smaller model that uses either the peak-to-peak value or the spectral moments. From the result in table 7.2 the Cp-value representing the geometric mean is larger than the value representing the peak-to-peak value. Because this method only looks at the importance of a specific covariate in the model, we only have an indication of the importance of the different terms and not of a specific model.

7.4 *Cp*-criterion on multiple models

The previous two methods used parameter selection from a single model, and in this way decided the significance of the different covariates in the model. Instead of a single model we will in this section consider different models, and see which one gives the best result. Instead of removing one variable from a model as in the last section, each model we want to test against is specified. This indicates the best model as regarded by the *Cp*-criterion. The nine models tested against are displayed below:

$$oc_i = \alpha_0 + \alpha_1 pktopk_i + \alpha_2 temp_i + \alpha_3 sal_i + \epsilon_i \tag{7.2}$$

$$oc_i = \alpha_0 + \alpha_1 m0_i + \alpha_4 temp_i + \alpha_5 sal_i + \epsilon_i \tag{7.3}$$

$$oc_i = \alpha_0 + \alpha_1 energy_i + \alpha_3 temp_i + \alpha_4 sal_i + \epsilon_i \tag{7.4}$$

$$oc_i = \alpha_0 + \alpha_1 geo_i + \alpha_2 temp_i + \alpha_3 sal_i + \epsilon_i \tag{7.5}$$

$$oc_i = \alpha_0 + \alpha_1 average_i + \alpha_2 temp_i + \alpha_3 sal_i + \epsilon_i \tag{7.6}$$

$$oc_i = \alpha_0 + \alpha_1 pktopk_i + \alpha_2 geo_i + \alpha_3 m1_i + \alpha_4 m2_i + \alpha_5 temp_i + \alpha_6 sal_i + \epsilon_i \tag{7.7}$$

$$oc_i = \alpha_0 + \alpha_1 m0_i + \alpha_2 m2_i + \alpha_3 temp_i + \alpha_4 sal_i + \epsilon_i \tag{7.8}$$

$$oc_i = \alpha_0 + \alpha_1 m0_i + \alpha_2 m2_i + \epsilon_i \tag{7.9}$$

and,

$$oc_i = \alpha_0 + \alpha_1 geo_i + \alpha_2 m2_i + \epsilon_i \tag{7.10}$$

The nine equations are choosed because they display interesting properties of the different predictors. At the same time we wish to keep the number of models at an acceptable level.

Model	Equation	Cp-value
$oc_i = \alpha_0 + \alpha_1 pktopk_i + \alpha_2 temp_i + \alpha_3 sal_i + \epsilon_i$	7.2	143482097
$oc_i = \alpha_0 + \alpha_1 m0_i + \alpha_2 temp_i + \alpha_3 sal_i$	7.3	115487509
$oc_i = \alpha_0 + \alpha_1 power_i + \alpha_2 temp_i + \alpha_3 sal_i + \epsilon_i$	7.4	116393179
$oc_i = \alpha_0 + \alpha_1 geo_i + \alpha_2 temp_i + \alpha_3 sal_i + \epsilon_i$	7.5	97333702
$oc_i = \alpha_0 + \alpha_1 average_i + \alpha_2 temp_i + \alpha_3 sal_i + \epsilon_i$	7.6	85529129
$oc_i = \alpha_0 + \alpha_1 pktopk_i + \alpha_2 geo_i + \alpha_3 m1_i + \alpha_4 m2_4 + \alpha_4 temp_i + \alpha_5 sal_i + \epsilon_i$	7.7	39255249
$oc_i = \alpha_0 + \alpha_1 m0_i + \alpha_2 m2_i + \alpha_3 temp_i + \alpha_4 sal_i + \epsilon_i$	7.8	63119910
$oc_i = \alpha_0 + \alpha_1 m0_i + \alpha_2 m2_i + \epsilon_i$	7.9	68205327
$oc_i = \alpha_0 + \alpha_1 geo_i + \alpha_2 m2_i + \epsilon_i$	7.10	50391809

Table 7.3: *Cp*-values produced by different models.

The variable $pktopk_i$ represents the peak-to-peak value, the variables $m0_i$, $m1_i$ and $m2_i$ correspond to the zero, first and second order moment. The variable geo_i uses the geometrical mean on the Fourier spectrum, and is related to the zero order moment as described in chapter 6. The variable $power_i$ uses the square sum of the elements in the filtered signal, as described in section 6.4.2. And the variable $average_i$ use the average value of the filtered signal to represent the photoacoustic response.

The strength of each equation is tested with the *Cp*-criterion and the result is found in table 7.3. The first five equations (7.2 to 7.6) are using five different methods as the photoacoustic response combining them with the variables temperature and salinity. From prior analyses the temperature and salinity are observed to be physical quantities that influence the photoacoustic response directly. It is believed that these two parameters are key values, and dangerous to remove in a final model. The first equation (7.2) uses the peak-to-peak value and is the method used today to find the photoacoustic response. It displays the highest Cp-value of 143482097 and therefore the worst result of the nine equations. Exchanging the peak-to-peak value with the zero order moment we get equation 7.3, with an improved Cp-value of 115487509. As showed in section 6.4.2 the zero order moment is related to the power of the signal through Parseval's theorem. This fact is also observed in equation 7.4 using the signal power, and as expected it receives a similar Cp-value to equation 7.3. The geometric mean used in the next equation is showing an improvement from the zero-order moment with a Cp-value of 97333702, proving that it is a good alternative. The best Cp-value of the five methods is displayed by equation 7.6, and uses the average of the filtered signal. This is both a surprising and very good result because this method at the same time is the easiest to equate. Another way to

approach the problem is to use all the variables when finding the oil concentration. This is done in equation 7.7. It is only based on the spectral moments, none of the values from the filtered signals are used. This method is the most complex to compute, but at the same time receives the best Cp-value of 39255249. We wanted to cut down on this model to see which variables have largest influence on the result. Equation 7.8 gives a good results based on the zero and second spectral moments, temperature and salinity. In equation 7.9 the temperature and salinity predictors are removed displaying a surprisingly good result. This result indicates that the spectral moments in themselves include enough information to find the oil concentration. This is a very interesting result, and should be researched further. It might be that this statement only is valid for temperatures and salinity within the small variations used in the test matrix. It would be unwise to trust blindly in the result from the statistics, and forget the dependencies showed earlier. It is advised that any model should include at least salinity and temperature. The last equations show that this result can be further improved by using the geometric mean instead of the zero order moment, and at the same time showing the strong properties of the geometric mean.

7.4.1 Residual plots of the models

Above nine different equations are presented. A low Cp-value indicates a good fit to the data. The value says nothing about the fitted values ability to model the true values. A residual plot manage to display such information. The Cp-value uses the residual sum of squares as a basis of computation, and a similar trend should be reflected in the residual plots. In this case the data is not normalized so the axes display correct scaling. The boxplots in this chapter are of the same type used in equation 5.7 and 5.8 in chapter 5. These plots use the peak-to-peak value processed by the instrument. This chapter uses the wavesignal dataset and the peak-to-peak value found from this. The plots in the two chapters should therefore not be compared directly.

The four plots in figure 7.2 to 7.5 show the residuals after regression performed on equations 7.2, 7.6, 7.7 and 7.10. Equation 5.3 in chapter 5 is identical to equation 7.2. This equation use the peak-to-peak value and corresponds to the method currently used in the instrument today. The residual plot are different, because the equations are working on two different datasets. The plot of figure 7.2 can therefore be viewed as the reference method. Figure 7.3 shows the residuals when using the average of the filtered signal. When comparing the residuals with the reference method we can observe that there is a small improvement in the prediction of the oil concentration, as expected from the Cp-values. The best Cp-value is found from equation 7.7 a method that includes

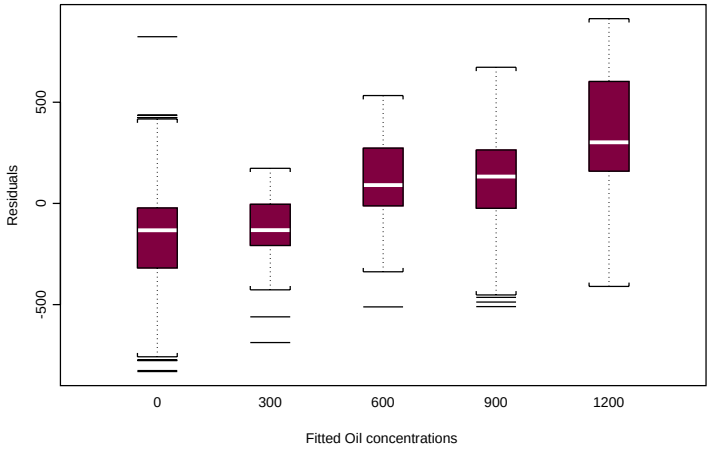

Figure 7.2: This plot shows the residuals of equation 7.2 after regression.

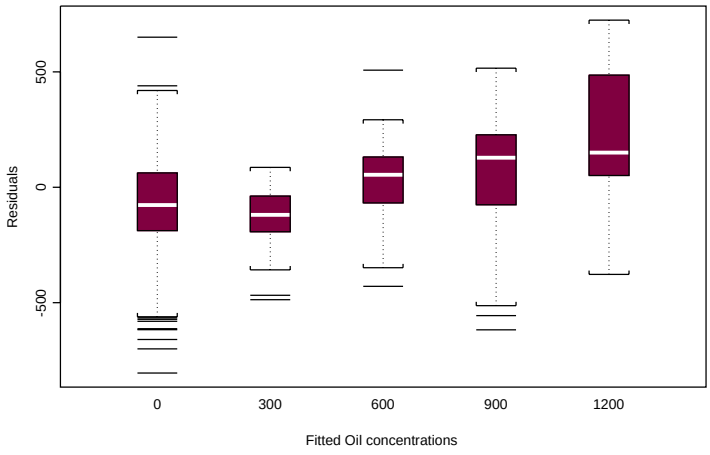

Figure 7.3: This plot shows the residuals of equation 7.5 after regression.

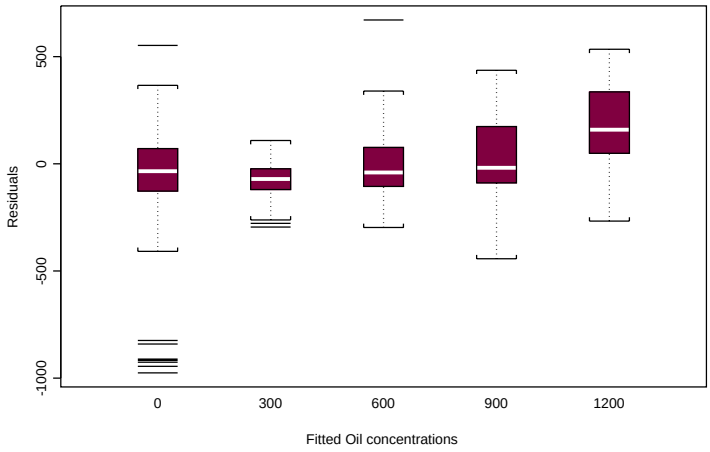

Figure 7.4: This plot shows the residuals of equation 7.7 after regression.

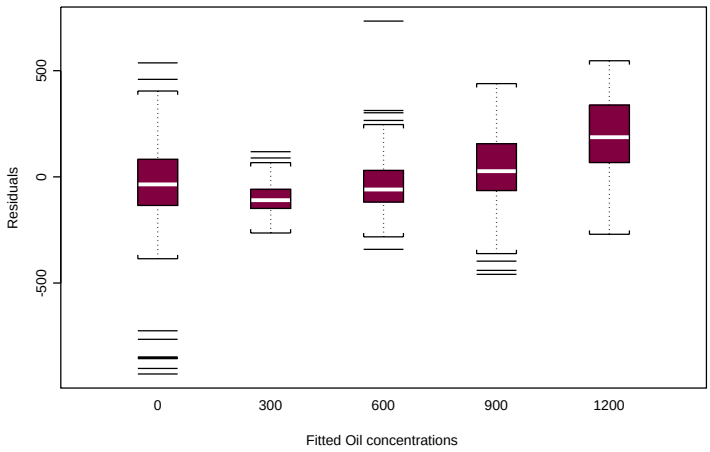

Figure 7.5: This plot shows the residuals of equation 7.10 after regression.

all the variables. This equation equals equation 7.1 used earlier in this chapter. The residual plot from this process is displayed in figure 7.4, the method displays a good fit, but is the gain in accuracy large enough to be worth the effort of extra processing. To illustrate the effect of removing the temperature and salinity predictors the residuals of equation 7.10 can be seen in figure 7.5. The method shows good ability to predict the oil concentration, and is surprisingly close in appearance to figure 7.4. It might be interesting to study this equation closer because of the advantage of being independent of salinity and temperature measurements. Especially the salinity where no good method of finding the value exists.

7.5 Discussion

From section 7.3 we can conclude that the geometric mean is the parameter that has strongest influence on the model of equation 7.1. This also indicates that the geometric mean has better ability to predict the oil concentration than the peak-to-peak value. This is the same result that we observed in section 7.4.

Models with interesting properties were constructed to explore a set of variables that will give a good result. In this case the optimal method could be to include each and all variables, both the spectral moments and the peak-to-peak value in the same model. The aim of this chapter was however to decide between one of the two. The models used when exploring the Cp-value indicate that the methods based on the spectral moments are better to predict the oil concentration. The difference in Cp-value tell us that the spectral moments are probably a very good estimator compared to the peak-to-peak value. From figure 7.1 we see that there is strong correlation between the zero order spectral moment and the peak-to-peak value. The second order spectral moment together with the geometric mean have good ability to predict oil concentration. From figure 7.1 we see that the second order moment have good correlation towards temperature, this might be the explanation for the good properties as a predictor. From the last analysis of multiple models, the Cp-value indicate that the spectral moments are sufficient in themselves to decide the oil concentration. Comparing the Cp-value of equation 7.3 and 7.9, and only using the second order moment yields a better result instead of using the temperature and salinity. Earlier observations have shown that the oil concentration is connected to physical factors such as temperature, pressure and salinity. Removing these components without more testing are dangerous, and will probably ruin the result. If the range of the test values are small, the influence of the components will not show up in the statistical analysis.

When exploring the residual plots the difference in performance indicated by the Cp-value is harder to discover. It is at this point no statistics on how much is gained from using one method compared to another. When choosing a method the performance versus the complexity of calculations should be considered. Equation 7.6 is a method that fulfill both criteria, from the Cp-value it has better performance than the peak-to-peak value, zero order moment, and geometric mean. It is also the method that is easiest to implement.

Chapter 8

Conclusion and further work

8.1 Conclusion

The analysis from chapter 5 has shown that it is possible to calibrate the current version of the instrument to a specific compound crude oil. In the oil concentration range of 0-750 ppm we observe that the interquartile range of the residual error is within ± 150 ppm for both the oil dispenser and the IR-method. The higher oil concentrations display higher error. It looks like the tendency is a gain in residual error at the higher levels of oil concentration. The main reason is believed to be an oil layer building up over time on the instrument window. This is coloured noise and affects all oil concentrations. The probe was withdrawn from the pipeline and cleaned three times during testing, the reason was that the oil concentration did not come down to the expected level when running normal sea water. If this noise is removed the accuracy in monitoring should be greatly enhanced. This issue must be solved before doing more testing, and is the main reason that the instrument did not have an acceptable performance.

Another problem is the oil dispenser, any error in the delivered oil concentration will directly affect the error in photoacoustic response. The IR-method has an error of 10 %, the performance of the oil dispensers is unknown. This uncertainty should be avoided in future analyses.

The two problems mentioned above are probably the main reason for much of the error in detection of oil concentration. By solving these problems the accuracy of the instrument will probably be improved within the specifications of ± 100 ppm set by Kværner.

The Fourier spectrum was inspected and the main components containing information were found to be laying in the frequency band 0.1-0.6 MHz. A filter should at least pass these frequencies. When filtering it is important that the filter has linear phase. Unlinear phase tend to

distort the energy in the signal, which has severe impact when evaluating the peak-to-peak value. This led to the use of spectral moments as an alternative method for equating the oil concentration. Only parts of the Fourier spectrum were used so that unwanted noise elements in the signal were removed from the calculations. The zero order spectral moment was the value that had greatest similarities to the peak-to-peak value, and from statistical analysis appears to be a better predictor. The effect of salinity on the photoacoustic response has not been properly tested. The salinity range used in the test matrix was too small to detect any finer changes. If this range is typical for salinity variations in produced water, there should not be any big reasons to calibrate the instrument to this value. It was found that the zero order spectral moment is connected to the mean square of the signal through Parseval's theorem. This way we can find the the zero order moment directly from the signal itself.

The choice of a model to find the photoacoustic response should be dependent on the usage of the instrument. The most accurate method is to use all available variables. Statistics show that most of the variables explored in this thesis have some influence on the oil concentration. Especially the spectral moments are suitable, most information exists in the zero order moment, but the higher order moments also have high explanatory properties. The model using the average of the filtered signal is also showing good properties, and is suitable in environments where there are reason to reduce the complexity of computations. It is also a nice method because it should be possible to omit the A/D card, which is appropriate in a downhole environment where we meet problems with high temperature and limited signal bandwidth.

The distance to the first peak was believed to contain information about the salinity concentration. A linear model was constructed to check if it was possible to detect salinity changes. The residuals from the process show a systematic error in the fitted values. The residual error fit the data within a $\pm$ 10 g/l, this is not good figures when the salinity were monitored in the range of 30-48 g/l. The systematic error indicates that an unlinear model might be more appropriate. From earlier analysis in laboratory [3] salinity dependencies were detected in the range from 0-100g/l, so the test range of salinity should be increased in further work to get a better base of data.

8.2 Further work

In this thesis it was shown the possibility of filtering the signal followed by the arithmetic mean to produce the photoacoustic response. This shows that it is possible to find the photoacoustic response with the aid

of a lowpass, highpass and an averaging filter. There still remain much work before an actual hardware implementation can be produced. Such an implementation should be of interest if the instrument is developed to work in a downhole environment. With the limited signal capacity to the surface, the photoacoustic response can be sent together with the other necessary parameters.

In chapter 6 it was shown that it was not necessary with the high sampling frequency of 100 MHz. This could be reduced down to 4 MHz when using the arithmetic mean of the filtered signal. Further testing showed that it was possible to reduce the sampling even lower without serious reduction in accuracy. A reasonable sampling frequency depends on which method will be used to equate the photoacoustic response. When using the method of averaging of the signal, the A/D card can possible be omitted. It should be mentioned that if the instrument is to be used to detect salinity, high sampling frequency is necessary to get an accurate value of the distance to the first peak.

More research should be performed with the spectral moments. The fact that the statistics claimed that the three first spectral moments were able to predict oil concentration better than models including both temperature and salinity is an interesting observation. It might be that calibration towards these values are unnecessary in future applications. On the other hand physical dependencies are known to exist. It is believed that the test range of salinity and temperature are to small to show any big influence on the oil concentration. It is possible that this influence has problems to appear through the noise in the observations.

The oilfilm that get stuck on the instrument window during monitoring must be solved. One solution is to produce some kind of reference monitoring. This was the reason the 1550 nm wavelength laser was added. It is possible that it can correct some of the error, but more research is needed. Another method could be to run pure water, at certain time intervals to monitor the effect of the oilfilm, then remove the extra response. The oil stuck to the window gives an added response in addition to the oil already in the water. It might be sufficient to move the pressure transducer further away from the window so the response from the oilfilm dissipate before reaching the sensor. This will not be possible if the oilfilm scatter the laser beam, resulting in less response instead. A last solution might be to add some kind of cleaning device, such as a high pressure or chemical washer.

8.2.1 Other applications

The Spectral moments should find use in other photoacoustic applications, and possible in other instrumentation using amplitude detection. The statistics did show correlation between the second order spectral

moment and the temperature. The fact that the zero order moment and the second order moment were able to predict the oil concentration better than methods including temperature and salinity indicate that the spectral moments contain additional information about the signal. It is possible that other applications using photoacoustics are able to find dependencies related to their physical parameters, and in this way gain new information. The arithmetic mean of the filtered signal will also be able to improve the measurements, it should also be an advantageous method in production of smaller and less expensive units.

References

[1] SFT-rapport 1762/2000, “Utslipp på Norsk kontinentalsokkel 1999, Olje kjemikalier og utslipp til luft”, 2000.

[2] United Kingdom ,“Prevention of Oil Pollution Act 1971’,HMSO.

[3] Scott Stuart Freeborn, “Pulsed Laser Photoacoustic instrumentation for the monitoring of crude oil in produced water”, Department of Physics, Herriot-Watt University, Edinburgh, September 1997.

[4] Allan Rosencwaig, “Photoacoustics and Photoacoustic Spectroscopy”, John Wiley & Sons 1980.

[5] Sian Ashton, “The Modelling and Application of Pulsed Laser Photoacoustic for the detection of body analytes” Department of Physics, Heriot-Watt University, Edinburgh, March 1999.

[6] Peter Hodgson, “The Detection of oil in water by near Infrared Pulsed photoacoustic Spectroscopy”Peter Hodgson,Department of Physics, Heriot-Watt University, Edinburgh, March 1994.

[7] John Hannigan, “Near Infrared Photoacoustic spectroscopy and Instrumentation for the detection of Hydrocarbons in Water”,Department of Physics, Heriot-Watt University, Edinburgh, March 1999.

[8] S. S. Freeborn, J. Hannigan, F. Greig, R. A. Suttie and H. A. MacKenzie, “A pulsed photoacoustic instrument for the detection of crude oil concentrations in produced water”, REVIEW OF SCIENTIFIC INSTRUMENTS, Vol 69, No. 11, November 1998

[9] F. Greig, E. M. Johnston, T. D. Binnie, H. A. MacKenzie, “A PC Based Photo-Acoustic Instrumentation System”, Ultrasonic Synposium 1994, page 1333-1336

[10] ”Water analysis:Determination of oil in water Infrared spectrophotometric method”,Norsk Standard NS9803, juni 1993.

[11] Tone Schanke, Scott S. Freeborn, “Oil In Water - Downhole Monitoring: Norsk Hydro Tests”, Kværner Oilfield Products, Heriot-Watt University, Document no. 407954, dec. 1998.

[12] Kværner Oilfield Products “OIWM PHASE 2 - FINAL REPORT” KOP specification No. 60-0ND0003-00, Date 2000.29.05.

[13] Kværner Oilfield Products “Oil In Water Monitor Characterisation Method” KOP specification No. 60-060045-06

[14] Data Analysis Products Division, Mathsoft, ,“S-PLUS 4 Guide to Statistics”, Seattle, July 1997.

[15] Ashish Sen, Muni Srivastava, “Regession Analysis: Theory, Methods and Applications”, Springer-Verlag, ISBN 0-387-97211-0, 1990

[16] Rafael C. Gonzales, Richard E. Woods “Digital Image Processing”, Addison Wesley Publishin Company, September 1993.

[17] Sverre Holm, IfI Memo, 13 august 1999

[18] T.Schanke, S.A.Kjølberg, F.Vogel, “Oil in Water Monitoring for Subsea and Downhole Separators”, SPE 66538 ,Society of Petrolium Engineers, February 2001

[19] A.V. Oppenheim, A.S. Willsky, S.H. Nawab, “Signals & Systems” second edition, prentice-hall international inc., ISBN 0-13-651175-9, 1997.

[20] A.V. Oppenheim, R.W. Schafer, “Discrete-Time Signal Processing”, prentice-hall international inc., ISBN 0-13-216771-9, 1989.

[21] Øyvind Hagen, Luren Yang, “Photoacoustic instrument for detection of crude oil concentrations in produced water”, Sintef 1999.

[22] Robert J. Urick, “Pinciples of underwater sound”, McGraw-Hill Book Company, ISBN 0-07-066087-5, 1983.

[23] V.A. Del Grosso, “equation for the speed of sound in natural waters (with comparison to other equations)”, The Journal of the Acoustic Society of America, may 1974.

[24] Jack R. Lowett, “Merged seawater sound-speed equations”, The Journal of the Acoustic Socity of America, sept 1977.

[25] P. Terzoudi, T. Whitaker, H.A. Mackenzie, “Use of Photo-acoustic measurement technology to measure hydrocarbon concentration levels in reinjection lines”, Kværner Oilfield Products Ltd., Heriot-Watt University, COSWASS JIP, 2000.

[26] The MathWorks, Inc., “www.mathworks.com”.

Appendix A

Matlab code used in filter and spectral moments design

```
% FIR filter design:

F=[0 0.003 0.006 0.015 0.02 1];
AFIR=[-4.3 1.9 1.2 1 0 0];
B=fir2(500, F, AFIR);
Y=conv(X,B);
figure(1)
freqz(B,1.0,512,1e8);

% IIR filter design:

[BIIR, C] = cheby2(3, 20, [0.0004 0.02]);
figure(2)
freqz(BIIR,C,512,1e8);
Z=filter(BIIR,C,X);


% Compare filter outputs
figure(3)
hold
plot((1:1000)/100,A)
plot((1:1000)/100,Y(245:1244)-2.8,'r')
plot((1:1000)/100,Z(1:1000)-1.8,'g')
xlabel('Time(ms)');
ylabel('Photoacoustic response');
grid

B=B-mean(B);
C=(N/C2)-C1;
```

```
A=abs(fft(B,N)).^2;
l=(1:N)-1;
k1=(l./(N*dt));

%Aritmetisk middel
Y0=(1/C)*sum(A(C1:round(N/C2)));

%Geometrisk middel
G=prod(A(C1:round(N/C2)).^(1/round(C)));

% normalize to m0 = 1 for the rest of the moments
A(C1:round(N/C2))=A(C1:round(N/C2))/Y0;
Y=(1/C)*sum(A(C1:round(N/C2)).*k1(C1:round(N/C2)));

% log to fit the log plot afterwards
X=Y0;

if n>=1,
  X=[X,Y];
end

for i = 2:n
  k2=((l./(N*dt))-Y).^i;
  Y2=(1/C)*sum(A(C1:round(N/C2)).*k2(C1:round(N/C2)));
  Y2=Y2^(1/i);  % from variance to standard deviation etc
  X=[X,Y2];
end
```

In chapter 6.4 it was used an alternative method to find the zero order spectral moment without Fourier transforming the signal. The Matlab code used in the examples is presented below.

```
function o=enkel(X)

%Initialising some variables.

Fs=100e6;          % sampling frequency
Ndes=25;           % decimation factor
Fs2=Fs/Ndes;       % sampling frequency after decimation
Nfir=41;           % Length of highpass filter

%Lowpass filter.
```

```
figure(1)
Nfir1=50;
F=[0  0.5*Fs2/2 0.8*Fs2/2 0.5*Fs]/(0.5*Fs);
M=[1     1     0     0 ];

B1=fir2(Nfir1,F,M,kaiser(Nfir1+1,0.5));
freqz(B1,1.0,512,Fs);

Y=conv(X,B1);

Y=Y(26:1026);

%decimation
Y2=Y([1:25:length(Y)-1]);

%Highpass filter.
figure(2)

B=-1/Nfir*ones(1,Nfir); % <=> trekke fra middel
B((Nfir+1)/2)=B((Nfir+1)/2)+1;
freqz(B,1.0,512,Fs2);

Y3=conv(Y2,B);

%Equating the arithmetic mean of the signal.

o=(sum(abs(Y3(21:60).^2)))*(1/39);
```

Appendix B

Report presented at the SPE/EPA/DOE Exploration and production Enviromental Conference

SPE 66538

Oil in Water Monitoring for Subsea and Downhole Separators

T.Schanke, SPE, Kvaerner Oilfield Products, S.A.Kjølberg, Norsk Hydro Research Center, and F.Vogel, University of Oslo

This paper was prepared for presentation at the SPE/EPA/DOE Exploration and Production Environmental Conference held in San Antonio, Texas, 26–28 February 2001.

Abstract

This paper describes the use of pulsed photoacoustics to measure ppm concentrations of oil in treated/separated water from subsea and downhole separators. The principle of operation and the system for the pulsed photoacoustic Oil-In-Water Monitor (OIWM) prototype is described. Main test results for the prototype sensor are presented, with focus on the impact on measurement by changes in temperature, pressure and salinity. Following , a model used to compensate such variations is shown. The paper finally addresses the implications of using pulsed photoacoustic measurement principle to measure oil entrainment in produced water from subsea and downhole separators.

Introduction

Significant quantities of produced water are recovered as a by-product of offshore production of gas and oil. The produced water is obtained from two main sources; formation water from the reservoir and seawater injected into the reservoir during oil and gas production. The produced water is either disposed to sea or re-injected into the reservoir.

A positive verification of the purity of the treated/separated water is an essential parameter for water treatment and separation plants. The industry has for some time evaluated various methods to find a potential method for further development into an in-line, continuos monitoring method for field applications top-side, sub-sea and downhole to replace the infrared spectrophotometric method[1] frequently used today, in this paper referred to as IR-analysis. This procedure describes a method where samples are taken manually for laboratory analysis. For subsea separation this implies the mobilisation of a ROV (Remote Operating Vehicle), which is a costly and difficult operation. The availability of a sensor for inline, continuos measurement of the water quality from subsea and downhole separators can drastically reduce the operational costs while improving the control of the separation process - which means less use of chemicals. A pulsed photoacoustic monitoring system can provide an attractive alternatives to today's method as it can be designed to operate inline to continuously give the oil concentration in produced water from topside, subsea and downhole separators. This method has today reached a prototype stage.

Principle of operation

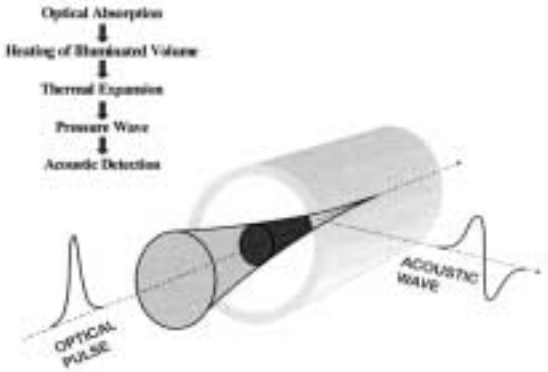

Figure 1: Principle of operation for pulsed photoacoustics

The Oil-In-Water Monitor (OIWM) system described in this article is based on pulsed photoacoustic principle[2-5]. The principle of operation is illustrated in Figure 1. Pulsed laser light is absorbed by the oil in the water stream. When the light is absorbed the optical energy is dissipated as heat which causes sudden local heating. The subsequent thermal expansion generates a pressure wave, which is detected with an acoustic detector. The amplitude of the acoustic peak is proportional to the oil concentration. Hence, by measuring the amplitude one can find the oil concentration.

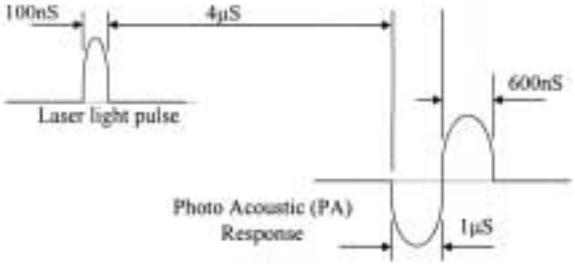

Figure 2: Oil In Water Monitor signals

OIWM system description

The OIWM prototype comprises of three main parts the Head Unit, the Control Unit and the User Interface PC, see Figure 3. A future subsea OIWM will consist of the Head Unit and the Control Unit, which will be directly linked to a standard subsea control pod. Re-design is needed for a downhole version of the OIWM.

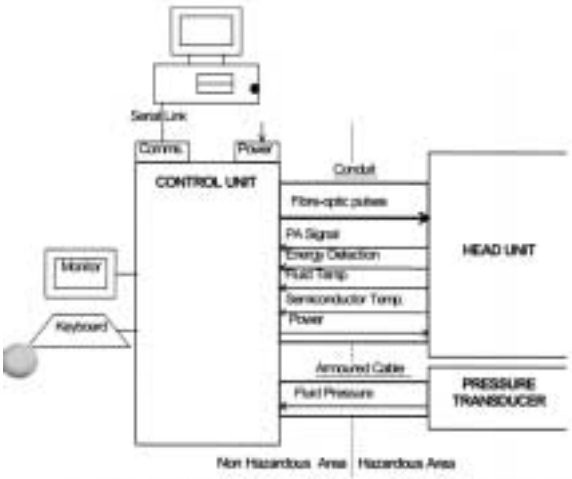

Figure 3: Oil In Water Monitor System Block Diagram[9]

Head Unit. The Head Unit consists of a housing rated to 100bar pipeline pressure. It is the interface into the flow and contains the laser delivery via a sapphire window and detection electronics for the photoacoustic response, as well as laser energy monitoring and temperature sensors. The head unit also includes a pressure transducer, see Figure 4.

Figure 4: Head Unit

Control Unit. The Control Unit contains the digitizing, data processing and logging electronics, as well as the laser diodes and drivers. The Control Unit is connected to the Head Unit via an armored cable through which the fiber optic and all other signals pass and one armored cable from the Pressure Transducer. The control unit sends out pulsed laser light and receives acoustic response, flow pressure and temperature. There are two laser diodes in the control unit giving out light at 905nm and 1550nm. A keyboard and monitor connected to the control unit allows direct control of the data acquisition and logging during test. The Control Unit operates independently, but to allow monitoring of the data during logging, some of the data is transferred to the User Interface PC via a serial link.

User Interface PC. The User Interface PC used only in the prototype system, is a standard PC, with custom software, which displays data from the Control Unit and provides indication of the system status.

Test results

Previous results. The application of the pulsed photoacoustic technique to detection of crude oil concentration in produced water has been investigated through various laboratory studies and through development of laboratory prototype instrumnetation[7-15], including field testing at Orkney Water Technology Center, Norsk Hydro Research Center and FRAMO's Test Facilities. The test proved that the pulsed photoacoustic technique can be used to measure oil concentration in produced water continuously and inline in the range of 0 – 2000ppm. Good linear crude oil trend was achieved during the tests for various crude oils, see Figure 5. The tests also proved that the principle is not effected by changes in flow rate or droplet size and that changes in the photoacoustic response is in agreement with theory for temperature, pressure and salinity.

Based on the good results with the laboratory prototype a development project was initiated to design and build a prototype and to develop a signal processing to allow continuos corrections for variations in pressure, temperature and salinity.

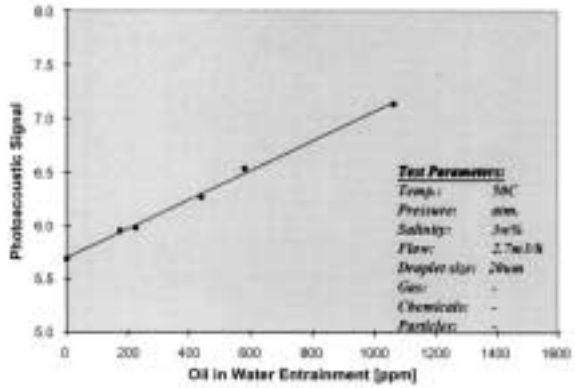

Figure 5: The graph shows the linear trend between oil entertainment in water and photoacoustic signal obtained with the lab prototype.

Test results with the OIWM prototype

Figure 6: OIWM site installation

During 1999 the OIWM prototype shown in Figure 3 was designed and two units were built. One of the OIWM was tested at Orkney Water Technology Centre[8] and the other at Norsk Hydro Research Center. This section of the paper presents the results from the tests at Norsk Hydro Research Center in Q1 2000.

The purpose of the test at Norsk Hydro Research Center was to obtain sufficient data to develop signal processing to give a ppm out-put continuously corrected for variations in temperature, pressure and salinity. The test parameters are given in Table 1.

Table 1: Test parameters

Test Parameter	Range	
Oil concentration	0, 300, 600, 900, 1200	ppm
Temperature	30, 35, 40, 45, 50	°C
Pressure	1,10, 20, 30	bar
Salinity	30, 38, 45, 48	g/l

Oil concentration. An oil dispenser set the oil concentration in the flow loop. The accuracy of the set point concentration is not known. It depends on the accuracy of the flow meter for the water and the oil dispenser. Samples were taken and IR-analyses were performed to measure the actual oil concentration achieved. The accuracy of neither the IR-oil concentration analyses nor the set-point ppm is known, however the correlation between the two methods was +/- 20%. A typical linear trend between oil in water entrainment given by IR-analysis and photoacoustic signal is plotted in Figure 5.

Temperature. The OIWM includes temperature sensor to measure local temperature during tests All test points taken during the test at Norsk Hydro Research Center with the prototype OIWM are plotted as a function of temperature in Figure 7. For low concentrations (0, 300, 600ppm) there is a clear trend in the signal. For higher concentrations (900 and 1200ppm) more problems both due to oil foaling of the sapphire window and accuracy in the reference method were observed.

Salinity. The test facility used a resistivity meter to control the salinity of the flow loop water. Earlier tests have shown that the photoacoustic response will increase with increasing salinity. Figure 8 shows that the variations due to salinity are relatively small compared to variations caused by changes in temperature or oil concentration.

Pressure. The OIWM includes pressure sensors to measure local pressure during tests. The OIWM was tested for pressures up to 30bar. No variations in the photoacoustic signal were observed as a function of pressure see Figure 9.

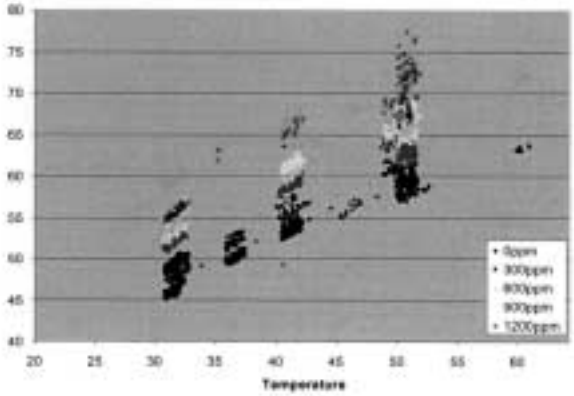

Figure 7: The graph shows all the test points taken during the test at Norsk Hydro Research Center. The photoacousitc response (PA 905nm) is plotted as a function of temperature for 0, 300, 600, 900 and 1200ppm.

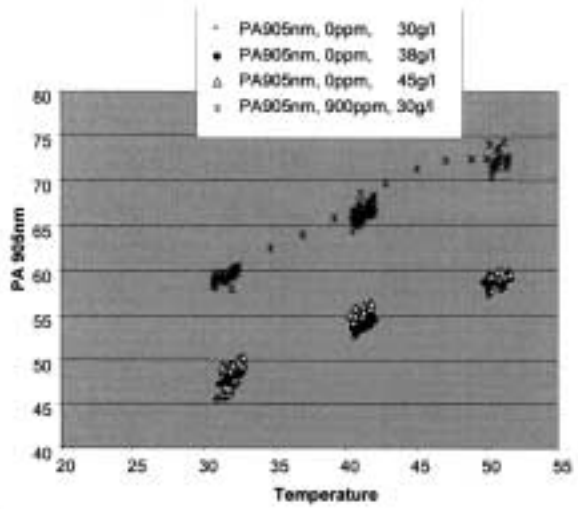

Figure 8: The graph shows the variations in the photoacoustic response caused by changes in salinity (30, 38 and 45g/l) compared to the variation caused by changes in oil concentration as a function of temperature.

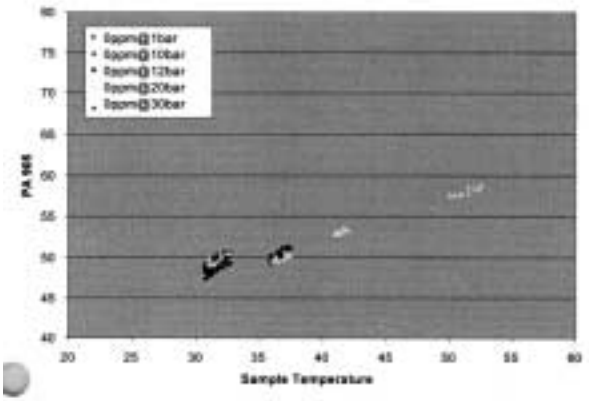

Figure 9: The graph shows how photoacoustic response for 905nm laser light depends on pressure.

Developing a model. The Photoacoustic response from the OIWM instrument is proportional to the oil content. For any given oil the oil concentration can be found by measuring the photoacoustic value. However, this value also varies with *temperature* (T), *pressure* (P) and *salinity* (S). The instrument should be capable of compensating for variations in these parameters during operation.

Multivariate Analyses (S-plus) was used on the test results presented in Figure 7 to find the relationship between the oil concentration (given by IR-analysis) and the instrument readings (T, P, PA).

Accurate and good reference values are important to get a good model, as any error from the reference method will be implemented in the model. There was poor correlation between the two reference methods, the oil set point and the IR-analyses. It is uncertain which of the two methods that are most accurate. However, a model corresponding to the IR-analysis is the most interesting, as this is the method currently used in the offshore industry.

Equation 1 gives the model related to the IR-analyses. How good the model matches the IR-analyses is displayed in Figure 10. The graph includes all combinations of temperature, pressure, salinity and oil concentration.

For every test point two samples were taken for IR-analyses. The repeatability in the IR value was +/-20% compared to the set point values, i.e. 300ppm +/-60ppm, 600ppm +/-120ppm, 900ppm +/- 180ppm and 1200ppm +/-240ppm. This explains some of the spread in Figure 10.

Build up of oil films was a major problem during the test and is likely to be another source for the spread in the results. The problem with oil films seemed to increase with increasing oil concentration.

$$ppm_{IR_i} = \alpha_0 + \alpha_1 PA_i + \alpha_2 T_i + \alpha_3 S_i \qquad (1)$$

Table 2 : Estimated parameters to equation (1). PA: photoacoustic response, T: Temperature, S: Salinity

Parameters	Values	Standard deviation
α_0	-1642,4	44,5
α_1	67,5	0,9
α_2	-40,6	0,8
α_3	-8,9	0,8

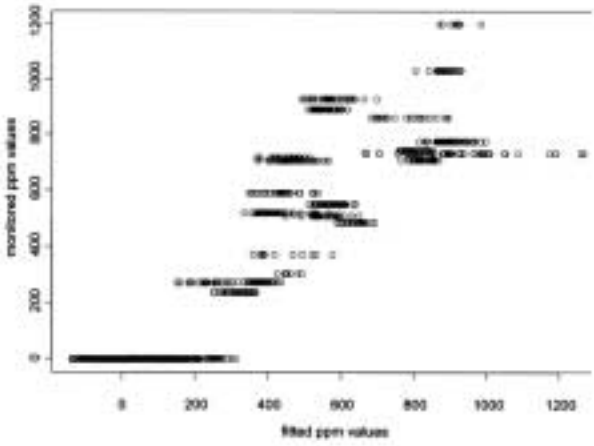

Figure 10: The graph shows the values obtained using the IR-model given by Equation 1 compared to the values measured by the reference method (IR-analyses).

Improving the model. The first step to improve the model is to prevent oil films and other type of build-ups on the sensor head. During the test at Norsk Hydro Research Center the flow velocity passed the head was relatively low, approximately 1m/s. A higher flow velocity is believed to reduce the problem of oil film build-ups. Secondly the method can be improved if the number of test points is increased and the control with the IR-results is improved. The last can be achieved by taking more samples for IR-analyses for each test point and use the average as the reference value. Furthermore the results can be improved allowing more time for each test point, for instant take the rolling average of the Photoacoustic signals during the last 15minutes. The current accuracy of the model compared to taking IR analyzes is +/-200ppm for the lower values (set point =0, 300, 600ppm). Implementing the above mentioned improvements a new model with an accuracy of +/- 100ppm, or even better, should be well within reach.

OIWM for Subsea and Downhole Separators

The first subsea separator world wide, Troll Pilot, was installed in May 2000 at the Troll field in the Norwegian sector of the North Sea[16]. Troll Pilot is, as the name indicates, a pilot project for subsea separation and injection. To control the quality of the injected water samples are taken by an ROV for lab analysis. This is a costly and difficult operation. As the water is re-injected this method is tolerable as no limitation is set by the government on the purity of the re-injected water. However, high concentration of oil in the re-injected water is an issue for the oil companies both from an economic and a technical point of view, as it might cause clogging of the reservoir. The availability of an inline photoacoustic OIWM can drastically reduce the operational costs. In addition it can offer continuos measurement of the water quality which can expand its use from a control instrument to also offer a method to improve the separation process. Improved control of the separator process will result in better control of the amount of chemical needed which indeed will be a benefit for the environment. For a subsea separator to be able to discharge the water to sea an inline water quality control is mandatory.

The first offshore installation of a downhole separator, H™-sep is planned for summer 2001. For a downhole separator the alternative to an inline method to control the water quality is even more complicated than for a subsea separator. For the pilot installation of the H™-sep a sample line to surface is planned. Work is ongoing to develop a downhole version of the pulsed photoacoustic OIWM to be integrated with H™-sep. This will dramatically improve the control of downhole separation.

Conclusion

An OIWM prototype based on pulsed photoacoustic has been built and tested. The system has proven to be stable and no drift over time has been observed. The instrument has been calibrated for Visund oil to detect oil concentrations in the range of 0-2000ppm, with automatic corrections for changes in temperature, pressure and salinity. It is set to displays ppm values corresponding to IR-analyses, which is the method currently used in the offshore industry. Inaccuracy in the oil concentration from the OIWM is mainly due to oil film built ups and the inaccuracy in the reference method. Improved model can only be installed once a mechanism to prevent oil films has been defined and tested and more test points with increased control of the IR-analyses has been performed. Increasing the flow velocity passed the instrument head is believed to be one method to reduce the oil film build-ups. Once these improvements have been performed a model with +/- 100ppm, or even better, should be well within reach.

For subsea and downhole separators a pulsed photoacoustic OIWM has the potential to offer inline, continuos control of the re-injected water, which can drastically reduce the operational costs while improving the control of the subsea and downhole separation process. The pulsed photoacoustic OIWM has reached a prototype stage.

Abbreviations

IR	Infrared
OIWM	Oil In Water Monitor
PA	Photoacoustic
ppm	part per million
ROV	Remote Operating Vehicle

Acknowledgement

The authors would like to express their gratitude to Norsk Hydro and The Norwegian Research Council who has funded this work. We would also like to thank Norsk Hydro Research Center for carrying out the test and Sverre Holm and Ole Christian Lingjarde from the University of Oslo for valuable discussion and help with the signal processing.

References

1. NS 9803. 1993. Determination of oil in water, Infrared spectrophotometric method.
2. Lai H M, Young K 1982 Theory of the pulsed photoacoustic technique *J. Acoust. Soc. Am.* **72** pp 2000-2007
3. Heritier J-M 1983 Electrostrictive limit and focusing effects in pulsed photoacoustic detection *Optics Comm.* **44** pp 267-272
4. Sullivan B, Tam A C 1984 Profile of laser-produced acoustic pulse in a liquid *J. Acoust. Soc. Am.* **75** pp 437-441
5. Tam A C 1986 Applications of photoacoustic sensing techniques *Rev. Mod. Phys.* **58** pp 381-431
6. P. Terzoudi, T. Whitaker, H.A. Mackenzie. 2000. Use of photo-acoustic measurement technology to measure hydrocarbon concentration levels in re-injection lines. North Sea Flow measurement Workshop.
7. Freeborn S S, Hannigan J, Greig F, Suttie R A, MacKenzie H A 1998 A pulsed photoacoustic instrument for the detection of

crude oil concentrations in produced water *Rev. Sci. Instrum.* **69** pp 3948-3952
8. Hannigan J, Greig F, Freeborn S S, MacKenzie H A 1999 A pulsed photoacoustic system for the spectroscopy and monitoring of hydrocarbon liquids using stimulated Raman scattering in a silica fibre as a near infrared source *Meas. Sci. Technol.* accepted for publication in February 1999 issue
9. Hodgson P, MacKenzie H A, Christison G B, Quan K-M Laser photoacoustic detection of organic analytes in aqueous media, in *Near infrared spectroscopy: bridging the gap between data analysis and NIT applications* (Ellis Horwood, 1992)
10. MacKenzie H A, Christison G B, Hodgson P, Blanc D 1993 A laser photoacoustic sensor for analyte detection in aqueous systems *Sensors and Actuators B* **11** p 213
11. Quan K-M, MacKenzie H A, Hodgson P, Christison G B 1993 Photoacoustic generation in liquids with low optical absorption *Ultrasonics* **32** p 181
12. Hand D P, Hodgson P, Carolan T A, Quan K-M, Mackenzie H A, Jones J D C 1993 Detection of Photoacoustic Waves in Liquids by Fibre Optic *Interferometry Opt. Comm.* **104** pp 1-6
13. Hodgson P, Quan K-M, MacKenzie H A, Freeborn S S, Hannigan J, Johnston E M, Greig F, Binnie T D 1994 Application of Pulsed Laser Photoacoustic Sensors in Monitoring Oil Contamination in Water *Sensors and Actuators B* **29** p 339
14. Hand D P, Freeborn S S, Hodgson P, Carolan T A, Quan K-M, Mackenzie H A, Jones J D C 1995 Optical Fibre Interferometry for Photoacoustic Spectroscopy in Liquids *Opt. Lett.* **20** pp 213-215
15. Greig F, Johnston E M, Binnie T D, MacKenzie H A, Freeborn S S 1994 A PC based photoacoustic instrumentation system *Proc. IEEE Ultrasonics Symp.* **3** pp 1333-6
16. http://www.hydro.no/